Inventiones mathematicae

Volume 53 1979

Springer-Verlag
Berlin · Heidelberg · New York

Inventiones mathematicae

Springer-Verlag Berlin · Heidelberg · New York

Printed in Germany by Universitätsdruckerei H. Stürtz AG, Würzburg

Index

Covered by Zentralblatt für Mathematik
and Contents of Mathematical Publications

Inventiones mathematicae

Vol. 53 Fasc. 1 1979

Covered by Zentralblatt für Mathematik and Contents of Mathematical Publications

Springer-Verlag
Berlin Heidelberg NewYork

Inventiones math. ISSN 0020-9910 INVMBH 53(1) 1–90 (1979)
24. VII. 1979

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

Inventiones math. 53, 1 – 44 (1979)

Inventiones mathematicae

Intersection Numbers of Sections of Elliptic Surfaces

David A. Cox and Steven Zucker*

Department of Mathematics, Rutgers University, New Brunswick, N.J. 08903, USA

The theory of elliptic surfaces over $\mathbf{C}$ draws on ideas and techniques from arithmetic, geometry and analysis. Let $\bar{f}$: $\bar{X} \to \bar{S}$ be a minimal elliptic fibration with non-constant j-invariant, which possesses a section σ_0. Then the group $\mathfrak{S}$ of sections of $\bar{f}$ can be naturally identified with the group of rational points of the generic fiber; i.e., $\mathfrak{S}$ consists of rational solutions of a cubic equation over a function field. On the analytic end, one can associate to these sections certain generalized automorphic forms which have a natural pairing (the Eichler bilinear form) defined on them. In this paper, we will examine the connection of these topics with the geometry of $\bar{X}$ via the intersection product and Hodge decomposition on its cohomology.

This paper had its origin in a problem posed by W. Hoyt: if the rank r of the group $\mathfrak{S}$ of sections is known (e.g., when $p_g = 0$) and one has r sections $\sigma_1, \ldots, \sigma_r$, do they form a basis of $\mathfrak{S}$ modulo torsion? The earlier attempt by Hoyt and Schwartz to answer this question involved a direct use of the Eichler pairing. Its values lie in $(1/N)\mathbf{Z}$ for some $N \in \mathbf{Z}$, and a bound for the best value of N can be computed from the monodromy. Thus, by looking at the discriminant of this form on the given sections, one obtains a sufficient (though not always necessary) condition for them to be generators modulo torsion. The hope had been that one would be able to compute this discriminant in examples. However, the calculations are very messy and have been done only in the simplest of cases. Another difficulty in this program is that sometimes one must base-change before the Eichler pairing can be defined.

We give an alternate method (using intersection numbers on $\bar{X}$) for doing such computations (§ 1), and we will use it to compute a large number of examples (§ 2). In § 3 we will show how our geometric methods relate to the automorphic forms and Eichler pairing, and we also determine the role they play in the Hodge theory of $\bar{X}$.

Now we discuss the contents of the paper in greater detail. Given $\bar{f}$: $\bar{X} \to \bar{S}$ as described above, let f: $X \to S$ be the smooth part of $\bar{f}$ (throw away the singular

* Supported in part by NSF Grant MCS 76-06364

0020-9910/79/0053/0001/$08.80

fibers), and let $j\colon S \to \bar{S}$ be the inclusion. In § 1 we define a homomorphism:

$$\delta\colon \mathfrak{S} \to H^1(\bar{S}, R^1\bar{f}_*\mathbf{Q}) \cong H^1(\bar{S}, j_* R^1 f_* \mathbf{Q})$$

as follows. Given $\sigma \in \mathfrak{S}$, we want to alter its cycle class $[\sigma] \in H^2(\bar{X}, \mathbf{Z})$ so that it lies in the first Leray filtration level L^1 (which consists of classes that restrict to zero on all of the fibers); in general, this can be done only if we pass to rational cohomology. The class $[\sigma - \sigma_0]$ gives us zero on all the good fibers, but if σ and σ_0 pass through different components of a singular fiber, some more corrections are needed. This is done by adding a *rational* linear combination D of components of singular fibers (and D is determined at each fiber solely by the component that σ hits – see (1.14)). Having now $\delta(\sigma) = [\sigma - \sigma_0 + D] \in L^1$, we define a bilinear pairing $\langle\ ,\ \rangle$ on $\mathfrak{S}$ by setting:

$$\langle \sigma, \sigma' \rangle = -(\delta(\sigma) \cup \delta(\sigma'))$$

for σ, σ' in $\mathfrak{S}$ (and the cup product is taken equivalently in $H^2(\bar{X}, \mathbf{Q})$ or in $H^1(\bar{S}, R^1\bar{f}_*\mathbf{Q})$). The pairing differs surprisingly little from the intersection number:

$$(0.1) \quad [\sigma - \sigma_0] \cdot [\sigma' - \sigma_0]$$

and the difference is again expressible in terms of the components of the bad fibers hit by the sections. The correction factors are listed in (1.19).

We have $\langle \sigma, \sigma \rangle \geqq 0$, and $\langle \sigma, \sigma \rangle = 0$ if and only if σ is torsion. The discriminant of $\langle\ ,\ \rangle$ on $\mathfrak{S}$ is easily computable in terms of the torsion of $\mathfrak{S}$, the singular fibers of $\bar{f}$, and the discriminant of intersection product on the Néron-Severi group $NS(\bar{X})$. Normally one does not know much about $NS(\bar{X})$, but when $p_g = 0$, $NS(\bar{X}) = H^2(\bar{X}, \mathbf{Z})$ and the discriminant of intersection product is 1 by Poincaré duality. Then sections $\sigma_1 \ldots \sigma_r$ of $\bar{f}$, where $r = \operatorname{rank} \mathfrak{S}$, are a basis of $\mathfrak{S}$ (modulo torsion) if and only if:

$$(0.2) \quad \det \langle \sigma_i, \sigma_j \rangle = (\#\mathfrak{S}_{\mathrm{tor}})^2 / \prod m_s$$

where m_s is the number of components of multiplicity one in the fiber $X_s = \bar{f}^{-1}(s)$. This formula is the basis for all of the examples in § 2. Then we examine the discriminant of cup product on $H^1(\bar{X}, R^1\bar{f}_*\mathbf{Z})$ and discuss (without proof) the relation between $\mathfrak{S}_{\mathrm{tor}}$ and $H^2(\bar{S}, R^1\bar{f}_*\mathbf{Z})$. The section ends with a treatment of the relation between $\langle\ ,\ \rangle$ and Tate heights.

In the Appendix to § 1 we consider *arbitrary* elliptic fibrations (possibly non-algebraic) $\bar{f}\colon \bar{X} \to \bar{S}$, and we determine sufficient (and often necessary) conditions for the maps $\bar{f}_*\colon \pi_1(\bar{X}) \to \pi_1(\bar{S})$ and $\bar{f}_*\colon H_1(\bar{X}, \mathbf{Z}) \to H_1(\bar{S}, \mathbf{Z})$ to be isomorphisms. In this paper we use these results to conclude that various cohomology groups are torsion-free.

§ 2 has a strong arithmetic flavor to it, since $\mathfrak{S}$ is isomorphic to the group of rational solutions of any Weierstrass equation:

$$(0.3) \quad y^2 = 4x^3 - g_2 x - g_3, \qquad g_2, g_3 \in K(\bar{S})$$

defining the generic fiber of $\bar{f}$. Typical of the examples we give is: the $\mathbf{C}(t)$-

rational solutions of $y^2=4x^3-3t^3x+t^4$ form an infinite cyclic group generated by $(0, t^2)$. All of the examples are surfaces with $p_g=0$ (so that (0.2) is true and $r=\operatorname{rank}\mathfrak{S}$ is known). Each example starts with explicit solutions $\sigma_1, \ldots, \sigma_r$ of an Eq. (0.3), which are shown to be a basis of $\mathfrak{S}$ (modulo torsion) by using (0.2). Most of the machinery for computing $\langle\sigma, \sigma'\rangle$ is developed in § 1 (see (1.18) and (1.19)), with two exceptions. First, we need to determine which component of a bad fiber is hit by a section. This requires looking at the bad fibers case by case (using the Kodaira classification) and it occupies a large part of § 2. Second, computing the intersection product (0.1) is complicated by the fact that all we have to work with is the Weierstrass equation, which does not define $\bar{X}$ at the singular fibers. So we must show how to avoid this difficulty. We also show how to compute $\mathfrak{S}_{\text{tor}}$, which is necessary for (0.2). Two of the examples are worked out in detail.

In § 3, we treat the relation between automorphic forms and the cohomology group $H^1(\bar{S}, j_* V)$, where $V=R^1 f_* \mathbf{C}$. We begin by observing that this group is naturally isomorphic to the first parabolic cohomology group associated to the monodromy representation of $\pi_1(S)$ on the first cohomology group of the fiber. A section $\sigma\in\mathfrak{S}$ gives rise to elements of $H^1(\bar{S}, j_* V)$ in three seemingly different ways:

1. The class $\delta(\sigma)$ from § 1.

2. Lifting σ to the upper half-plane $\mathfrak{h}$ (the universal cover of S), σ becomes expressible as an analytic function $F\colon \mathfrak{h}\to\mathbf{C}$, which possesses a period cocycle.

3. Let $\tau\colon \mathfrak{h}\to\mathfrak{h}$ denote the period function for X/S, lifted to $\mathfrak{h}$. Then one obtains a generalized automorphic form of weight 3 (3/2 to some) from F above by differentiating twice with respect to τ. Such automorphic forms feed naturally into the parabolic cohomology.

It is not hard to see that in fact all three are equal ((3.9) and (3.10)). Using this, we prove (3.12) that the generalized Eichler pairing for the automorphic forms coming from sections σ and σ' is none other than our pairing $\langle\sigma, \sigma'\rangle$. Consequently, we are working with the same pairing as Hoyt and Schwartz, with a more efficient way of evaluating it.

We conclude (§ 3C) with results on the Hodge decomposition of $H^1(\bar{S}, j_* V)$. By [25], there is a filtered complex comprised of locally-free $\mathcal{O}_{\bar{S}}$-modules (extending $\Omega_S^{\bullet}(V)$) whose hyper-cohomology gives the cohomology of $j_* V$, such that the induced filtration on cohomology gives a Hodge structure of weight two:

$$H^1(\bar{S}, j_* V)=H^{2,0}\oplus H^{1,1}\oplus H^{0,2};$$

moreover, it coincides with the one induced by $H^2(\bar{X}, \mathbf{C})$ through the Leray spectral sequence. Using this, we show ((3.20) and (3.24)) that filtration levels $F^2=H^{2,0}$ and $F^1=H^{2,0}\oplus H^{1,1}$ are naturally isomorphic to certain spaces of automorphic forms. The description of F^1 involves the ramification divisor of τ on $\bar{S}$. In this, we must make a reasonable extension to $\Sigma=\bar{S}-S$ of the notion of ramification (3.17). Then $\dim H^{1,1}$ is equal to the total order of ramification of τ, which yields a clearer proof (3.21) of a result in Shioda's paper [23]. Similar in spirit, we reprove in (3.22) another proposition from [23], giving an upper bound on the rank of $\mathfrak{S}$.

We are grateful to W. Hoyt for getting us interested in the problem described above. He conjectured some of the results proved in this paper. The idea of using a correction term consisting of a rational linear combination of components of bad fibers was suggested by P. Deligne. We would also like to thank G. Winters and R. Miranda for several useful conversations.

§ 1. The Pairing on Sections

An elliptic fibration over $\mathbf{C}$ is a map $\bar{f}$: $\bar{X} \to \bar{S}$ where $\bar{S}$ is a smooth compact curve over $\mathbf{C}$, $\bar{X}$ is a smooth compact surface over $\mathbf{C}$, and the generic fiber of $\bar{f}$ is an elliptic curve. We will assume that $\bar{f}$ has a section σ_0 and that there are no exceptional curves of the first kind in the fibers of $\bar{f}$; then $\bar{f}$: $\bar{X} \to \bar{S}$ is the Néron model of its generic fiber (and $\bar{f}$ is what Kodaira calls the "basic member"); see [14] and [8, § 8]. We will also assume that the j-invariant of $\bar{f}$ is non-constant.

The fiber $\bar{f}^{-1}(s)$, $s \in \bar{S}$, will be denoted X_s, and we let $S = \{s \in \bar{S}, X_s$ is a nonsingular fiber$\}$. The set $\Sigma = \bar{S} - S$ is finite, and the X_s, $s \in \Sigma$ are the "bad fibers" of $\bar{f}$. We will use Kodaira's classification of the bad fibers into types I_b $(b > 0)$, I_b^* $(b \geqq 0)$, II, II*, III, III*, IV, and IV* (see [8, § 6]).

Let X be the inverse image of S in $\bar{X}$, so that we have a commutative diagram:

$$(1.1) \qquad \begin{array}{ccc} X & \hookrightarrow & \bar{X} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle \bar{f}} \\ S & \xhookrightarrow{j} & \bar{S} \end{array}$$

where f is proper and smooth. For an abelian group M, the locally constant sheaf $V_M = R^1 f_* M$ on S corresponds to the monodromy representation of $\pi_1(S)$ on $H^1(X_t, M)$ (where $t \in S$). M will be either $\mathbf{Z}, \mathbf{Q}$ or $\mathbf{C}$, and $V_{\mathbf{C}}$ will be denoted simply V. By abuse of notation we will sometimes write $V_M = H^1(X_t, M)$.

Finally we recall the group structure of the fibers. Each good fiber X_t, $t \in S$, becomes a group with $\sigma_0(t)$ as the identity. For bad fibers, things are more complicated: given $s \in \Sigma$, take the union of the components of multiplicity one of X_s, and delete all singular points. We call this X'_s, and it becomes a group with $\sigma_0(s)$ as the identity (X'_s is an extension of $\mathbf{C}$ or $\mathbf{C}^*$ by the finite group of components of multiplicity one – see [14, III.17]). The component of X_s containing $\sigma_0(s)$ is called the zero component.

Then $\bar{X}' = X \cup \bigcup_{s \in \Sigma} X'_s$ is a commutative group variety over $\bar{S}$ with σ_0 as zero section. Since any section of $\bar{f}$: $\bar{X} \to \bar{S}$ lands in $\bar{X}'$, the set $\mathfrak{S}$ of all sections of $\bar{f}$ is a group with σ_0 as the identity. Since the j-invariant of $\bar{f}$ is non-constant, $\mathfrak{S}$ is finitely generated (this is the Mordell-Weil Theorem). Let $\mathfrak{S}_0$ be the subgroup consisting of those sections which hit the zero component of X_s for $s \in \Sigma$.

The above notation will be used throughout the paper.

A. We first compute the map $R^1\bar{f}_*\mathbf{Q}\to j_* R^1 f_*\mathbf{Q}$:

(1.2) **Lemma.** *The map* $R^1\bar{f}_*\mathbf{Q}\to j_* R^1 f_*\mathbf{Q}+j_* V_{\mathbf{Q}}$ *is an isomorphism.*

Proof. Let T be the local monodromy transformation for $s\in\Sigma$. Then the map $(R^1\bar{f}_*\mathbf{Q})_s\to(j_* R^1 f_*\mathbf{Q})_s$ becomes a map:

(1.3) $H^1(X_s, \mathbf{Q})\to H^1(X_t, \mathbf{Q})^T$

where X_t is a good fiber for some t near s. Using [8, §6 and §9], one sees that the groups in (1.3) are either both 0 or both $\mathbf{Q}$. Since (1.3) is a surjection by the local invariant cycle theorem (see [1]), it must be an isomorphism. □

It is well known that $(V_{\mathbf{Z}})^{\pi_1(S)}$ is zero and $H^2(\bar{S}, j_* R^1 f_*\mathbf{Z})$ is finite (see [8, §11]). Thus, from (1.2), we see that $H^0(\bar{S}, R^1\bar{f}_*\mathbf{Q})=H^2(\bar{S}, R^1\bar{f}_*\mathbf{Q})=0$. From this we easily see that all differentials in the Leray spectral sequence for $\bar{f}$ over $\mathbf{Q}$ vanish. Hence:

(1.4) **Lemma.** *The Leray spectral sequence of* $\bar{f}$: $\bar{X}\to\bar{S}$ *degenerates at* E_2 *over* $\mathbf{Q}$. □

This is also true for quite general reasons – see [25, §15].

Let us make some remarks about the situation over $\mathbf{Z}$. With a little care, one can improve (1.2) to show that $R^1\bar{f}_*\mathbf{Z}\to j_* V_{\mathbf{Z}}$ is an isomorphism. Then $H^2(\bar{S}, R^1\bar{f}_*\mathbf{Z})$ is finite, yet $H^3(X, \mathbf{Z})$ is torsion-free by (1.48). From this one easily proves that the differential:

(1.5) $d_2^{0,2}$: $H^0(\bar{S}, R^2\bar{f}_*\mathbf{Z})\to H^2(\bar{S}, R^1\bar{f}_*\mathbf{Z})$

is surjective, and that the Leray spectral sequence degenerates at E_2 over $\mathbf{Z}$ if and only if $H^2(\bar{S}, R^1\bar{f}_*\mathbf{Z})=0$. This is interesting in light of (1.30).

One can also prove that $H^1(\bar{S}, R^1\bar{f}_*\mathbf{Z})$ is torsion-free, a result of [23].

B. We need some notation. The Leray filtration on $H^2(\bar{X}, \mathbf{Q})$ is $L^2\subseteq L^1\subseteq L^0=H^2(\bar{X}, \mathbf{Q})$, where, by (1.4):

$$L^1=\ker(H^2(\bar{X}, \mathbf{Q})\to H^0(\bar{S}, R^2\bar{f}_*\mathbf{Q})),$$
$$L^1/L^2\simeq H^1(\bar{S}, R^1\bar{f}_*\mathbf{Q}),$$
$$L^2=\operatorname{im}(H^2(\bar{S}, \mathbf{Q})\to H^2(\bar{X}, \mathbf{Q}))=\mathbf{Q}\cdot[X_t] \quad (t\in S).$$

Also, for $s\in\Sigma$, we write $X_s=\sum_{i\geqq 0} m_i^s C_i^s$, where we label the C_i^s so that C_0^s is the zero component of X_s.

(1.6) **Theorem.** *Let* σ *be in* $\mathfrak{S}$. *Then:*

1. *There is a rational linear combination* $\sum_{s\in\Sigma} D_s$ *of the components of bad fibers* ($D_s=\sum_i a_i^s C_i^s$, $a_i^s\in\mathbf{Q}$) *so that* $[\sigma-\sigma_0+\sum_s D_s]$ *lies in* L^1.

2. *The cohomology class* $[\sigma-\sigma_0+\sum_{s\in\Sigma} D_s]$ *gives a well-defined element* $\delta(\sigma)$ *of* $H^1(\bar{S}, R^1\bar{f}_*\mathbf{Q})$, *and the* map $\delta\colon \mathfrak{S}\to H^1(\bar{S}, R^1\bar{f}_*\mathbf{Q})$ *is a homomorphism.*

3. *Each* D_s, $s\in\Sigma$, *is unique up to a rational multiple of* X_s, *and is computed as follows. Assume that* σ *hits* C_k^s. *If* $k=0$, *we can choose* $D_s=0$; *if* $k\neq 0$, *then* D_s *satisfies the equations:*

$$(1.7)\quad D_s\cdot C_i^s=\begin{cases} 1 & i=0\\ -1 & i=k\\ 0 & \textit{otherwise.}\end{cases}$$

Proof. L^1 consists of those elements which restrict to zero in every fiber. For any irreducible curve C on $\bar{X}$, there is a commutative diagram:

$$(1.8)\quad \begin{tikzcd} H^2(\bar{X},\mathbf{Q}) \arrow[r] \arrow[dr] & \mathbf{Q} \\ & H^2(C,\mathbf{Q}). \arrow[u, "\wr"'] \end{tikzcd}$$

Then one easily shows that $a\in L^1$ if and only if $a\cdot C_i^s=0$ for all s and i. Thus, the assertion that $[\sigma-\sigma_0+\sum_s D]\in L^1$ is equivalent to the statement:

$$(1.9)\quad (\sigma_0-\sigma)\cdot C_i^s=D_s\cdot C_i^s \qquad \text{for all } s \text{ and } i.$$

Note that $D_s\cdot X_s=(\sigma_0-\sigma)\cdot X_s=0$, so that

$$D_s\cdot C_0^s=-\sum_{i>0} m_i^s D_s\cdot C_i^s \quad\text{and}\quad (\sigma_0-\sigma)\cdot C_0^s=-\sum_{i>0} m_i^s(\sigma_0-\sigma)\cdot C_i^s.$$

If we set $D'_s=D_s-a_0^s X_s=\sum_{i>0} b_i^s C_i^s$, we then see that (1.9) is equivalent to:

$$(1.10)\quad (\sigma_0-\sigma)\cdot C_i^s=D'_s\cdot C_i^s \qquad \text{for all } s \text{ and } i>0.$$

Since the matrix $(C_i^s\cdot C_j^s)_{i,j>0}$ is negative definite (see [23, Lemma 1.3]), (1.10) has a unique solutions. When $k=0$, $D'_s=0$, so D_s is a multiple of X_s, and when $k\neq 0$, the unique solution D'_s of (1.10) gives a solution D_s of (1.7) that is unique up to a multiple of X_s.

Since $L^2=\mathbf{Q}\cdot[X_t]$ and $[X_t]=[X_s]$, the uniqueness above shows that $\delta(\sigma)=[\sigma-\sigma_0+\sum_{s\in\Sigma} D_s]$ is a well-defined element of $L^1/L^2=H^1(\bar{S}, R^1\bar{f}_*\mathbf{Q})=H^1(\bar{S}, j_* V_{\mathbf{Q}})$. To show that δ is a homomorphism, we need only check that the map:

$$(1.11)\quad \mathfrak{S}\to H^1(\bar{S}, j_* V_{\mathbf{Q}})\to H^1(S, V_{\mathbf{Q}})$$

is a homomorphism. (The Leray spectral sequence of j yields that $H^1(\bar{S}, j_* V_{\mathbf{Q}})\to H^1(S, V_{\mathbf{Q}})$ is injective.) The map (1.11) clearly sends σ to the cohomology class $[\sigma-\sigma_0]$, and the proof of Proposition 3.9 of [24] shows that this map is a homomorphism. $\square$

(1.12) *Remark.* In §3 we observe that $\delta(\sigma)$ lies in the (1, 1) part of the natural Hodge structure on $H^1(\bar{S}, R^1\bar{f}_*\mathbf{C}) = H^1(\bar{S}, j_* V)$ constructed in [25].

Take $\sigma \in \mathfrak{S}$. We can normalize the D_s described in (1.6) so that $D_s \cdot \sigma_0 = 0$. This defines D_s uniquely, so we write it $D_s(\sigma)$. Since $D_s(\sigma)$ is actually determined by the component (necessarily of multiplicity one) of X_s hit by σ, it is easy to compute all of the possibilities.

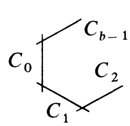
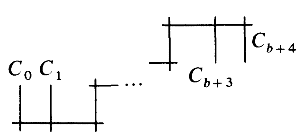
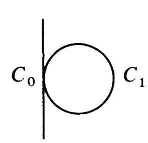
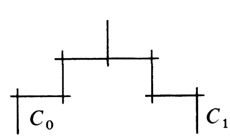
We first need to explicitly describe the bad fibers X_s. We use Kodaira's classification [8, §6] (see (1.13) below).

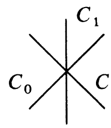
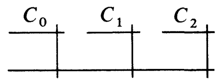
We have labeled only the components of multiplicity one; they are all we need for computations (see (1.19) and §2). We also drop the superscript s, and fiber types II and II* have been omitted because they both have only C_0 as a component of multiplicity one.

From (1.13) it is easy to find $D_s(\sigma)$ which satisfies (1.7) and $D_s(\sigma) \cdot \sigma_0 = 0$. The results are listed in (1.14) below.

C. The homomorphism δ of (1.6) will enable us to use cup product on $H^1(\bar{S}, R^1\bar{f}_*\mathbf{Q})$ to get a pairing $\langle\ ,\ \rangle$ on $\mathfrak{S}$. We will see below that $\langle\ ,\ \rangle$ has several nice properties, and in §2 it will play a crucial role (via (1.26)) in determining when we have generators (modulo torsion) of the group of sections of an elliptic surface with $p_g = 0$.

(1.13) Structure of the bad fibers X_s

Type	Structure	Picture
I_b $(b>0)$	$C_0 + C_1 + \ldots + C_{b-1}$	
I_b^* $(b \geqq 0)$	$C_0 + C_1 + 2C_2 + \ldots + 2C_{b+2} + C_{b+3} + C_{b+4}$	
III	$C_0 + C_1$	
III*	$C_0 + C_1 + 2C_2 + 2C_3 + 3C_4 + 3C_5 + 4C_6 + 2C_7$	

IV	$C_0+C_1+C_2$	
IV*	$C_0+C_1+C_2$ $+3C_3+2C_4+2C_5+2C_6$	

(1.14) Table for finding $D_s(\sigma)$

Type of X_s	Component hit by σ	$D_s(\sigma)$
Arbitrary	C_0	0
I_b	C_k $0<k<b$	$(b-k)/b\cdot C_1+2(b-k)/b\cdot C_2+\dots+k(b-k)/b\cdot C_k$ $+k(b-k-1)/b\cdot C_{k+1}+\dots+k/b\cdot C_{b-1}$
I_b^*	C_1	$C_1+C_2+\dots+C_{b+2}+1/2\cdot C_{b+3}+1/2\cdot C_{b+4}$
	C_{b+3}	$1/2\cdot C_1+2/2\cdot C_2+3/2\cdot C_3+\dots+(b+2)/2\cdot C_{b+2}+(b+4)/4\cdot C_{b+3}$ $+(b+2)/4\cdot C_{b+4}$
	C_{b+4}	$1/2\cdot C_1+2/2\cdot C_2+3/2\cdot C_3+\dots+(b+2)/2\cdot C_{b+2}+(b+2)/4\cdot C_{b+3}$ $+(b+4)/4\cdot C_{b+4}$
III	C_1	$1/2\cdot C_1$
III*	C_1	$3/2\cdot C_1+C_2+3/2\cdot C_3+2C_4+5/2\cdot C_5+3C_6+3/2\cdot C_7$
IV	C_1	$2/3\cdot C_1+1/3\cdot C_2$
	C_2	$1/3\cdot C_1+2/3\cdot C_2$
IV*	C_1	$4/3\cdot C_1+2/3\cdot C_2+C_3+5/3\cdot C_4+4/3\cdot C_5+2C_6$
	C_2	$2/3\cdot C_1+4/3\cdot C_2+C_3+4/3\cdot C_4+5/3\cdot C_5+2C_6$

The whole Leray spectral sequence for $\bar{f}\colon \bar{X}\to\bar{S}$ has cup products; in particular, there is a cup product

$$\cup\colon H^1(\bar{S},R^1\bar{f}_*\mathbf{Q})\otimes H^1(\bar{S},R^1\bar{f}_*\mathbf{Q})\to\mathbf{Q}$$

compatible with the usual cup product on $H^2(\bar{\mathrm{X}},\mathbf{Q})$. Then, for σ and σ' in $\mathfrak{S}$, we define:

$$\langle\sigma,\sigma'\rangle=-(\delta(\sigma)\cup\delta(\sigma')).$$

(1.15) **Lemma.** *$\langle\ ,\ \rangle$ is a bilinear form on $\mathfrak{S}$. Furthermore, for $\sigma\in\mathfrak{S}$, $\langle\sigma,\sigma\rangle\geqq 0$, and $\langle\sigma,\sigma\rangle=0$ if and only if σ is torsion.*

Proof. $\langle\ ,\ \rangle$ is bilinear because δ is a homomorphism. If σ is in $\mathfrak{S}$, choose an integer n so that $n\sigma$ is contained in $\mathfrak{S}_0$, i.e., $n\sigma$ hits the zero component of X_s for all $s\in\Sigma$. Then (1.14) show that $\delta(n\sigma)=[n\sigma-\sigma_0]$. Thus

$$n^2\langle\sigma,\sigma\rangle=\langle n\sigma, n\sigma\rangle=-(n\sigma)^2+2(n\sigma)\cdot\sigma_0-\sigma_0^2.$$

But all sections have the same self-intersection σ_0^2, which is a negative number by (2.4). So $n^2\langle\sigma,\sigma\rangle=-2\sigma_0^2+2(n\sigma)\cdot\sigma_0$. If σ is not torsion, then $n\sigma$ and σ_0 are distinct divisors, so that $(n\sigma)\cdot\sigma_0\geqq 0$, which implies $n^2\langle\sigma,\sigma\rangle>0$. The lemma follows. □

(1.16) **Corollary.** *δ: $\mathfrak{S}\to H^1(\bar{S}, R^1\bar{f}_*\mathbf{Q})$ is injective modulo torsion.* □

The proof of (1.15) also yields the following known fact (see [23, Prop. 1.6]):

(1.17) **Corollary.** *$\mathfrak{S}_0$ is torsion-free.* □

Since cup product is negative definite on the (1, 1) part of the Hodge structure on $H^1(\bar{S}, j_*V)$, we get another proof (via (1.12)) that $\langle\sigma,\sigma\rangle\geqq 0$ for $\sigma\in\mathfrak{S}$.

The first step in computing $\langle\sigma,\sigma'\rangle$ is:

(1.18) **Lemma.** *For σ,σ' in $\mathfrak{S}$, $\sigma\cdot D_s(\sigma')=\sigma'\cdot D_s(\sigma)$, and*

$$\langle\sigma,\sigma'\rangle=-(\sigma-\sigma_0)\cdot(\sigma'-\sigma_0)-\sum_{s\in\Sigma}\sigma\cdot D_s(\sigma').$$

Proof. Suppose that σ hits C_k^s, and write $D_s(\sigma')=\sum_{i>0} a_i\cdot C_i^s$. Then $\sigma\cdot D_s(\sigma')=a_k$, and from (1.7) we see that $D_s(\sigma)\cdot D_s(\sigma')=-a_k$. The first equality of the lemma then follows by symmetry, and the second is now an easy computation. □

Thus $\langle\sigma,\sigma'\rangle$ has a "geometric part," $-(\sigma-\sigma_0)\cdot(\sigma'-\sigma_0)$, and then "correction terms" coming from the behavior of σ and σ' at the bad fibers. Using (1.14), it is easy to make a table giving all the correction terms:

(1.19) Table of the local correction terms $\sigma\cdot D_s(\sigma')$

Type of X_s	Criterion (see (1.13))	Correction Factor
Arbitrary	σ or σ' hits the zero component	0
I_b	σ hits C_k, σ' hits $C_{k'}$, $0<k\leqq k'$	$k(b-k')/b$
I_b^*	σ,σ' hit C_1	1
	σ,σ' both hit C_{b+3} or C_{b+4}	$(b+4)/4$
	One hits C_{b+3}, the other C_{b+4}	$(b+2)/2$
	One hits C_1, the other C_{b+3} or C_{b+4}	1/2
III	σ,σ' hit C_1	1/2
III*	σ,σ' hit C_1	3/2

IV	σ, σ' both hit C_1 or C_2	2/3
	One hits C_1, the other hits C_2	1/3
IV*	σ, σ' both hit C_1 or C_2	4/3
	One hits C_1, the other hits C_2	2/3

Note that the greatest denominator that can occur is the *exponent* of the group G_s of components of X_s of multiplicity one (see [14, III.17] – type I_b^* is especially interesting). This tells us the following:

(1.20) **Corollary.** *Let N be the l.c.m. of the exponents of the groups of components of multiplicity one of the bad fibers X_s, $s \in \Sigma$. Then $N\langle\ ,\ \rangle \in \mathbf{Z}$.* □

Finally, we want to compare $\langle\ ,\ \rangle$ on $\mathfrak{S}$ to the usual intersection form $(\ ,\ )$ on $NS(\bar{X})$, the Néron-Severi group of $\bar{X}$ (note that by (1.48), $NS(\bar{X})$ is torsion-free). Forms like these have a discriminant defined as follows. Let $(\ ,\ )$ be a $\mathbf{Q}$-valued form on a finitely generated abelian group G of rank r. If $\sigma_1, \ldots, \sigma_r$ generate G modulo torsion, then define:

(1.21) $\operatorname{disc}(\ ,\ )_G = \det(\sigma_i, \sigma_j)/(\#G_{\mathrm{tor}})^2$

where G_{tor} is the torsion subgroup of G (and the determinant of a 0×0 matrix is 1). Then we have:

(1.22) **Proposition.** *Let m_s be the number of components of multiplicity one in the fiber X_s. Then:*

$$\operatorname{disc}\langle\ ,\ \rangle_{\mathfrak{S}} = |\operatorname{disc}(\ ,\ )_{NS(\bar{X})}| / \prod_{s\in\Sigma} m_s.$$

Proof. If $\sigma = \sum b_i \sigma_i$ in $\mathfrak{S}$, then the divisors $\sigma - \sigma_0$ and $\sum b_i(\sigma_i - \sigma_0)$ are linearly equivalent on the generic fiber of $\bar{f}$ (Abel's theorem). Thus, in $NS(\bar{X})$ we have:

(1.23) $\sigma - \sigma_0 = \sum b_i(\sigma_i - \sigma_0) + aX_t + \sum_{i>0} b_i^s C_i^s$

where X_t and the C_i^s are as in (1.6). From this and Theorem 1.1 of [23], we see that the map sending σ to $\sigma - \sigma_0$ gives an isomorphism:

(1.24) $\mathfrak{S} \simeq NS(\bar{X})/(\mathbf{Z}[\sigma_0] + \mathbf{Z}[F] + \sum_{i>0} \mathbf{Z}[C_i^s])$.

Let H be the subgroup of $NS(\bar{X})$ generated by the classes of σ_0, X_t, C_i^s (for $i>0$) and σ (for $\sigma \in \mathfrak{S}_0$). If $\sigma_1, \ldots, \sigma_r$ is a basis of $\mathfrak{S}_0$, then (1.23) shows that H is spanned by $[\sigma_0]$, $[X_t]$, $[C_i^s]$ ($i>0$) and $a_j = [\sigma_j - \sigma_0 - ((\sigma_j - \sigma_0)\cdot\sigma_0)X_t]$ (note that $a_j \cdot C_i^s = a_j \cdot X_t = a_j \cdot \sigma_0 = 0$). Then one easily computes that:

(1.25) $|\operatorname{disc}(\ ,\ )_H| = \operatorname{disc}\langle\ ,\ \rangle_{\mathfrak{S}_0} \cdot \prod_s m_s$

because $\det(a_i, a_j) = \det\langle\sigma_i, \sigma_j\rangle$ and $\det(C_i^s \cdot C_j^s) = m_s$ (see Lemma 1.3 in [23]).

Let G_s be the group of components of multiplicity one in X_s. The natural map $\mathfrak{S} \to \oplus G_s$ (evaluating which components a section hits) gives us, via (1.24), a homomorphism $NS(\bar{X}) \to \oplus G_s$. This kernel of this map is H (use (1.24)), so that $[NS(\bar{X})\colon H] = [\mathfrak{S}\colon \mathfrak{S}_0]$, and then (1.25) and Lemma 1.8 of [23] give us:

$$\begin{aligned}\operatorname{disc}\langle\ ,\ \rangle_{\mathfrak{S}} &= \operatorname{disc}\langle\ ,\ \rangle_{\mathfrak{S}_0}/[\mathfrak{S}\colon \mathfrak{S}_0]^2 \\ &= |\operatorname{disc}(\ ,\ )_H|/[\mathfrak{S}\colon \mathfrak{S}_0]^2 \cdot \prod_{s\in\Sigma} m_s \\ &= |\operatorname{disc}(\ ,\ )_{NS(\bar{X})}|/\prod_{s\in\Sigma} m_s. \quad \square\end{aligned}$$

This proposition is proved (under a very restrictive hypothesis) in [23, Corollary 1.7].

If $p_g(\bar{X}) = 0$, then $NS(\bar{X}) = H^2(\bar{X}, \mathbf{Z})$ and $(\ ,\ )$ has discriminant 1 by Poincaré duality. From (1.21) and (1.22) we then get:

(1.26) **Corollary.** *If* $p_g(\bar{X}) = 0$, *then* $\operatorname{disc}\langle\ ,\ \rangle_{\mathfrak{S}} = 1/\prod_{s\in\Sigma} m_s$, *and this means the following:*

1. *If* $\mathfrak{S}$ *has rank* 0, *then* $(\#\mathfrak{S})^2 = \prod_{s\in\Sigma} m_s$.
2. *If* $\mathfrak{S}$ *has rank* $r > 0$, *then* $\sigma_1, \ldots, \sigma_r$ *in* $\mathfrak{S}$ *generate modulo torsion if and only if:*

$$\det\langle\sigma_i, \sigma_j\rangle = (\#\mathfrak{S}_{\mathrm{tor}})^2/\prod_{s\in\Sigma} m_s. \quad \square$$

We can also compute the discriminant of cup product on $H^1(\bar{S}, R^1\bar{f}_*\mathbf{Z})$:

(1.27) **Proposition.**

$$\operatorname{disc}(\ ,\ )_{H^1(\bar{S}, R^1\bar{f}_*\mathbf{Z})} = \left(\prod_{s\in\Sigma} m_s\right)/(\# H^2(\bar{S}, R^1\bar{f}_*\mathbf{Z}))^2.$$

Proof. We will just sketch the proof. Let L^1 be the kernel of $\pi\colon H^2(\bar{X}, \mathbf{Z}) \to H^0(\bar{S}, R^2\bar{f}_*\mathbf{Z})$, so that $L^1/\mathbf{Z}[X_t] \simeq H^1(\bar{S}, R^1\bar{f}_*\mathbf{Z})$. Let N be generated by L^1, σ_0 and the C_i^s. Then the proof of (1.22) (especially (1.25)) shows that:

$$|\operatorname{disc}(\ ,\ )_N| = |\operatorname{disc}(\ ,\ )_{H^1(\bar{S}, R^1\bar{f}_*\mathbf{Z})}| \cdot \prod_{s\in\Sigma} m_s.$$

Thus, we need to determine the index of N in $H^2(\bar{X}, \mathbf{Z})$, which is the same as the index of their images under the map π above.

Using (1.8) and Lemma 1.3 of [23], one easily sees that the image of N in $H^2(X_s, \mathbf{Z})$ has index m_s, so that $\pi(N)$ has index $\prod_{s\in\Sigma} m_s$ in $H^0(\bar{S}, R^2\bar{f}_*\mathbf{Z})$. Since $\pi(H^2(\bar{X}, \mathbf{Z}))$ has index $\# H^2(\bar{S}, R^1\bar{f}_*\mathbf{Z})$ (this is (1.5)), we are done. $\square$

(1.28) **Corollary.** $(\# H^2(\bar{S}, R^1\bar{f}_*\mathbf{Z}))^2 \,\Big|\, \prod_{s\in\Sigma} m_s$. $\square$

We can say some more about the relation of $\# H^2(\bar{S}, R^1\bar{f}_*\mathbf{Z})$ to other invariants.

(1.29) **Proposition.**

$$[\mathfrak{S}:\mathfrak{S}_0]\cdot(\#H^2(\bar{S},R^1\bar{f}_*\mathbf{Z}))\Big|\prod_{s\in\Sigma}m_s.$$

If $p_g(\bar{X})=0$, *then:*

1. $[\mathfrak{S}:\mathfrak{S}_0]\cdot(\#H^2(\bar{S},R^1\bar{f}_*\mathbf{Z}))=\prod_{s\in\Sigma}m_s$.
2. $\#H^2(\bar{S},R^1\bar{f}_*\mathbf{Z})\Big|[\mathfrak{S}:\mathfrak{S}_0]$.
3. *If* $\mathfrak{S}$ *is finite, then*

$$\#\mathfrak{S}=\#H^2(\bar{S},R^1\bar{f}_*\mathbf{Z})=(\prod_{s\in\Sigma}m_s)^{1/2}.$$

Proof. Since $\pi(N)\subseteq\pi(NS(\bar{X}))\subseteq\pi(H^2(\bar{X},\mathbf{Z}))$, we see that

$$[\pi(NS(\bar{X})):\pi(N)]\cdot(\#H^2(\bar{S},R^1\bar{f}_*\mathbf{Z}))\Big|\prod_{s\in\Sigma}m_s,$$

with equality when $p_g=0$. If H is as in the proof of (1.22) (where we discovered that $[NS(\bar{X}):H]=[\mathfrak{S}:\mathfrak{S}_0]$), then we see that $\pi(N)=\pi(H)$. Thus, we need only prove that the kernel of $NS(\bar{X})\to H^0(\bar{S},R^2\bar{f}_*\mathbf{Z})$ lies in H. This follows from (1.6), (1.24) and that fact that $D_s(\sigma)$ has integral coefficients only if it is zero (see (1.14)). □

The following is also true:

(1.30) **Proposition.** *There are natural isomorphisms:*

$$\mathfrak{S}_{\mathrm{tor}}\simeq H^1(S,R^1f_*\mathbf{Z})_{\mathrm{tor}}\simeq\mathrm{Hom}(H^2(\bar{S},R^1\bar{f}_*\mathbf{Z}),\mathbf{Q}/\mathbf{Z}).$$

Thus $\#\mathfrak{S}_{\mathrm{tor}}=\#H^2(\bar{S},R^1\bar{f}_*\mathbf{Z})$ (cf. (1.27)–(1.29)).

[We had originally conjectured a version of (1.30) for the case $p_g=0$. Subsequently, P. Deligne showed that (1.30) is true in general, and A. Kas later gave a different proof. The most natural proof (not given here) uses the isomorphism

$$\mathfrak{S}_{\mathrm{tor}}\xrightarrow{\sim}H^1(S,R^1f_*\mathbf{Z})_{\mathrm{tor}}$$

and Poincaré duality.]

D. Let E be the generic fiber of $\bar{f}:\bar{X}\to\bar{S}$ (so that $\mathfrak{S}=E(K)$, where $K=K(\bar{S})$), and let D be a divisor on E. Then, as described in [13], there is a height function h_D: $\mathfrak{S}\to\mathbf{Z}$ which measures the "size" of rational points on E. Since E is an abelian variety over K, there is a unique quadratic function $\hat{h}_D$: $\mathfrak{S}\to\mathbf{R}$ (the Tate height relative to D – see [13]) such that:

(1.31) $h_D=\hat{h}_D+O(1)$.

$\hat{h}_D$ being quadratic means that we can write it as

$$\hat{h}_D(\sigma)=(1/2)\,f(\sigma,\sigma)+\ell(\sigma),$$

where f (resp. ℓ) is bilinear (resp. linear) on $\mathfrak{S}$.

We want to describe $\hat{h}_D$ in terms of $\langle\ ,\ \rangle$. We first lift $\delta\colon \mathfrak{S}\to H^1(\bar{S}, R^1\bar{f}_*\mathbf{Q})$ of (1.6) to a homomorphism mapping into $H^2(\bar{X}, \mathbf{Q})$:

(1.32) **Lemma.** *The map* $\delta\colon \mathfrak{S}\to H^2(\bar{X}, \mathbf{Q})$ *defined by*

$$\delta(\sigma)=[\sigma-\sigma_0+\sum_s D_s(\sigma)-((\sigma-\sigma_0)\cdot\sigma_0)\,X_t]$$

(X_t *is a good fiber*) *is a homomorphism.*

Proof. $\delta(\sigma)$, as defined above, is the unique element of L^1 which satisfies $\delta(\sigma)\cdot\sigma_0=0$ and reduces to the $\delta(\sigma)$ of (1.14) in $H^1(\bar{S}, R^1\bar{f}_*\mathbf{Q})$. It follows easily that δ is a homomorphism. □

The divisor D on E gives a unique divisor $\bar{D}$ on $\bar{X}$ which contains no component of any fiber and pulls back to D on E. Then the Tate height h_D can be expressed in terms of $\bar{D}$, $\langle\ ,\ \rangle$, and δ as follows:

(1.33) **Theorem.** *For* $\sigma\in\mathfrak{S}$, *we have:*

$$\begin{aligned}\hat{h}_D(\sigma)&=(1/2)\langle\sigma,\sigma\rangle\deg D+(\bar{D}\cdot\delta(\sigma))\\&=\bar{D}\cdot(\sigma-\sigma_0+\sum_s D_s(\sigma)-(1/2)(\sum_s\sigma\cdot D_s(\sigma))\,X_t).\end{aligned}$$

Proof. The two formulas on the right hand side are equal by (1.32) and (1.18) (note that $\deg D=\bar{D}\cdot X_t$), and they define a quadratic function which we call $g(\sigma)$. Since we have

$$h_D(\sigma)=\bar{D}\cdot\sigma$$

(see [13, Theorem 4]), we get the formula:

$$h_D(\sigma)-g(\sigma)=\deg D-\sum_s\bar{D}\cdot D_s(\sigma)+1/2(\sum_s\sigma\cdot D_s(\sigma))\deg D.$$

Using (1.14) and (1.19), it is easy to find a constant C (depending only on $\bar{D}$ and the bad fibers) such that

$$|h_D(\sigma)-g(\sigma)|\leqq C$$

for all $\sigma\in\mathfrak{S}$. Since $\hat{h}_D(\sigma)$ is the unique quadratic function with this property, we must have $\hat{h}_D=g$. □

We can use this to strengthen a result of [13]:

(1.34) **Corollary.** *The following are equivalent:*

1. $\sigma\in\mathfrak{S}_0$.
2. *For every divisor* D *on* E, $\hat{h}_D(\sigma)=\bar{D}\cdot(\sigma-\sigma_0)$.

Proof. 1 ⇒ 2. If $\sigma\in\mathfrak{S}_0$, then $D_s(\sigma)=0$ for every $s\in\Sigma$, and the second statement follows from (1.33).

2 ⇒ 1. Let $D=\sigma$. Then we get

$$\hat{h}_D(\sigma)-\bar{D}\cdot(\sigma-\sigma_0)=(1/2)\sum_s\sigma\cdot D_s(\sigma),$$

and (1.19) shows that the left hand side of the above is zero only when $\sigma \in \mathfrak{S}_0$. □

We can also describe how our methods compare to those used by Néron in [15]. The basic tool used in [15] is a pairing $(D, \mathfrak{A})$ defined for divisors D and $\mathfrak{A}$ on E, where $\deg \mathfrak{A} = 0$. This gives a bilinear pairing:

$$\langle \sigma, \sigma' \rangle_D = (D_\sigma - D, \sigma' - \sigma_0).$$

(1.35) **Proposition.** *Let σ, σ' be in $\mathfrak{S}$. Then:*

1. $(D, \sigma - \sigma_0) = -(1/2) \langle \sigma, \sigma \rangle \deg D - (\bar{D} \cdot \delta(\sigma))$,
2. $\langle \sigma, \sigma' \rangle_D = -\langle \sigma, \sigma' \rangle \deg D$.

Proof. This follows from (1.33) and Proposition 11 of [15, II.14] (where Néron shows that $\hat{h}_D(\sigma) = -(D, \sigma - \sigma_0)$ with $-(1/2)\langle \sigma, \sigma \rangle_D$ as its bilinear part). □

From (1.33) and (1.35) we get

$$(D, \sigma - \sigma_0) = -\bar{D} \cdot (\sigma - \sigma_0) - \sum_s [(\bar{D} \cdot D_s(\sigma)) - (1/2)(\sigma \cdot D_s(\sigma)) \deg D],$$

which, when compared to the formula (see [15, III.6])

$$(D, \sigma - \sigma_0) = -\bar{D} \cdot (\sigma - \sigma_0) + \sum_s j_s(D, \sigma - \sigma_0),$$

leads us to conjecture that, for $s \in \Sigma$, we have:

$$j_s(D, \sigma - \sigma_0) = -(\bar{D} \cdot D_s(\sigma)) + (1/2)(\sigma \cdot D_s(\sigma)) \deg D.$$

Appendix to § 1

The First Homotopy and Homology Groups of an Elliptic Surface

Let $\bar{f}: \bar{X} \to \bar{S}$ be an arbitrary elliptic fibration (not necessarily algebraic). Then $\bar{f}$ induces maps $\bar{f}_*: \pi_1(\bar{X}) \to \pi_1(\bar{S})$ and $\bar{f}_*: H_1(\bar{X}, \mathbf{Z}) \to H_1(\bar{S}, \mathbf{Z})$, and we want to know when these maps are isomorphisms. For the fundamental group, we give necessary and sufficient conditions for this to be true (see (1.36)). For homology, our results are not as complete (see (1.40), (1.44) and (1.47)).

We will say that $\bar{f}$ has non-trivial local monodromy if there is $s \in \Sigma$ so that $R^1 \bar{f}_* \mathbf{Z}$ is non-constant in a neighborhood of s. This is equivalent to the minimal model of $\bar{f}$ having at least one fiber not of type ${}_mI_0$, $m \geqq 0$. If the j-invariant of $\bar{f}$ is non-constant, this condition is certainly satisfied.

(1.36) **Proposition.** *Let $\bar{f}: \bar{X} \to \bar{S}$ be an elliptic fibration, and let $\bar{S}$ have genus g. The following are equivalent:*

1. *$\bar{f}_*: \pi_1(\bar{X}) \to \pi_1(\bar{S})$ is an isomorphism.*
2. *$\bar{f}$ has non-trivial local monodromy, and:*

a) *If $g \geqq 1$, then $\bar{f}$ has no multiple singular fibers.*

b) *If $g = 0$, then $\bar{f}$ has $\leqq 2$ multiple singular fibers, and if it does have 2, then their multiplicities are relatively prime.*

Proof. $1 \Rightarrow 2$. We first consider the multiple singular fibers of $\bar{f}$. We can assume that $\bar{f}$ is minimal (this does not affect π_1). If $g = 0$, then $\pi_1(\bar{X}) = 0$, and we are done by the proof of Proposition 2 in [9]. If $g \geqq 1$, we will show that the existence of one multiple singular fiber X_s (of multiplicity m) leads to a contradiction.

Let $S_0 = \bar{S} - \{s\}$. Because $g \geqq 1$, we can find a finite ramified normal covering $\bar{S}' \to \bar{S}$ with group G, unramified over S_0, where every preimage of s has ramification index m. If $\bar{X}'$ is the normalization of $\bar{X} \times_{\bar{S}} \bar{S}'$, then $\bar{X}' \to \bar{S}'$ is an elliptic fibration, but more importantly, the map $\bar{X}' \to \bar{X}$ is a covering space with group G (see [9, § 1]).

The isomorphism $\pi_1(\bar{X}) \xrightarrow{\sim} \pi_1(\bar{S})$ means that every covering space of $\bar{X}$ is the pull-back of a covering space of $\bar{S}$. Yet the covering $\bar{X}' \to \bar{X}$ constructed above clearly cannot arise in this manner. Thus, we reach a contradiction.

Next, assume that $\bar{f}$ has trivial local monodromy. If $g \geqq 1$, then $\bar{f}$ has no multiple singular fibers, so that $\bar{f}$ must be smooth. We then get an exact sequence:

$$1 \to \pi_1(X_t) \to \pi_1(\bar{X}) \xrightarrow{\bar{f}_*} \pi_1(\bar{S}) \to 1$$

since $\pi_2(\bar{S}) = 0$. This is impossible because $\bar{f}_*$ is an isomorphism. When $g = 0$, assume that $\bar{f}$ is minimal, so that the only bad fibers are of type ${}_mI_0$, $m > 0$. Then Theorems 6 and 7 of [20, Ch. IV] show that $\chi(\bar{X}) = 0$, so that $q = p_g + 1 \geqq 1$ by Noether's formula. Since $\bar{f}_*$ is an isomorphism, $b_1 = 0$. Then $q = 0$ by [10, Theorem 3], again giving us a contradiction.

The proof of $2 \Rightarrow 1$ is an immediate consequence of the following two lemmas.

(1.37) **Lemma.** *Let g be the genus of $\bar{S}$, and assume the following:*

1. *If $g \geqq 1$, then $\bar{f}$ has no multiple singular fibers.*

2. *If $g = 0$, then $\bar{f}$ has $\leqq 2$ multiple singular fibers, and if it does have 2, then their multiplicities are relatively prime.*

Then, for any good fiber X_t, we have an exact sequence:

$$\pi_1(X_t) \to \pi_1(\bar{X}) \to \pi_1(\bar{S}) \to 1.$$

Proof. The bad fibers of $\bar{f}$ are X_s, $s \in \Sigma$ where now each one has multiplicity m_s, and as usual $S = \bar{S} - \Sigma$. Taking the fundamental groups of (1.1) gives a commutative diagram:

$$\begin{tikzcd}
 & & 1 & 1 & \\
1 \arrow[r] & K_3 \arrow[r] & \pi_1(\bar{X}) \arrow[r] \arrow[u] & \pi_1(\bar{S}) \arrow[r] \arrow[u] & 1 \\
 & \pi_1(X_t) \arrow[r] \arrow[u] & \pi_1(X) \arrow[r] \arrow[u] & \pi_1(S) \arrow[r] \arrow[u] & 1 \\
 & & K_2 \arrow[r] \arrow[u] & K_1 \arrow[u] & \\
 & & 1 \arrow[u] & 1 \arrow[u] &
\end{tikzcd} \tag{1.38}$$

where K_1, K_2 and K_3 are the appropriate kernels. The second row is exact because f is a C^∞-fibration, and the columns and first row are exact because $\pi_1(X)\to\pi_1(\bar{X})$ and $\pi_1(S)\to\pi_1(\bar{S})$ are onto.

Let C be a component of multiplicity n in X_s, and let u be a loop around it in X. Then u is in K_2, and if g_s is a loop around s in S, then u maps to g_s^n in K_1. If $\bar{f}$ has no multiple singular fibers, then we can assume that C has multiplicity one, so that u maps to g_s. Thus, $K_2\to K_1$ is onto, and then an easy diagram chase shows that $\pi_1(X_t)\to K_3$ is onto (which is precisely what we want to prove).

If $g=0$, then K_1 is the group generated by the g_s, $s\in\Sigma$, subject to the single relation $\prod_{s\in\Sigma} g_s=1$. If we have only one multiple singular fiber, say X_s, then the image of K_2 contains $g_{s'}$ for all $s'\neq s$. Since these generate K_1, we again conclude that $\pi_1(X_t)\to K_3$ is onto. If we have two multiple singular fibers, X_{s_1} and X_{s_2}, then the image of K_2 contains $g_{s_1}^{m_{s_1}}$, $g_{s_2}^{m_{s_2}}$ and $g_{s'}$ for $s'\neq s_1$ or s_2. Since m_{s_1} and m_{s_2} are relatively prime, the image is again all of K_1. □

(1.39) **Lemma.** *Assume that $\bar{f}$ has non-trivial local monodromy. Then the map $\pi_1(X_t)\to\pi_1(\bar{X})$ is zero for any good fiber X_t.*

Proof. First, assume that the j-invariant of $\bar{f}$ is constant. We can assume that $\bar{f}$ is minimal (blowing down does nor affect π_1). Then the non-trivial monodromy of $\bar{f}$ must come from a fiber X_s of type I_0^*, II, II*, III, III*, IV or IV*, all of which are simply connected. If Δ is a small disc around s in $\bar{S}$, then $\pi_1(X_t)\to\pi_1(\bar{X})$ factors:

$$\pi_1(X_t)\to\pi_1(\bar{f}^{-1}(\Delta))\to\pi_1(\bar{X}).$$

Since $\bar{f}^{-1}(\Delta)$ has the same homotopy type as X_s, $\pi_1(\bar{f}^{-1}(\Delta))=0$, so that $\pi_1(X_t)\to\pi_1(\bar{X})$ factors through zero.

Next, assume that the j-invariant is non-constant.

If $\bar{S}'\to\bar{S}$ is a map of curves, let $\bar{X}'$ be a resolution of singularities of $\bar{X}\times_{\bar{S}}\bar{S}'$. Then we have a commutative diagram:

$$\begin{tikzcd} \bar{X}' \arrow[r] \arrow[d, "\bar{f}'"'] & \bar{X} \arrow[d, "\bar{f}"] \\ \bar{S}' \arrow[r] & \bar{S} \end{tikzcd}$$

which shows that $\pi_1(X_t)\to\pi_1(\bar{X})$ factors through $\pi_1(X_t)\to\pi_1(\bar{X}')$. So we need only show that the latter map is zero. As we saw above, we can assume $\bar{X}'$ is minimal over $\bar{S}'$.

Find $\bar{S}'$ so that $\bar{X}'$ has no multiple singular fibers. Then $\bar{X}'\to\bar{S}'$ is a deformation of an algebraic elliptic surface since j is non-constant (see [8, § 11]). This does not change the map $\pi_1(X_t)\to\pi_1(\bar{X}')$, so we can assume that $\bar{X}$ is algebraic. Pulling back further if necessary, we can assume that the generic fiber of $\bar{f}$ has an affine equation $y^2=x(x-1)(x-\tau)$, where $\tau\in K(S)$ is non-constant (because j is).

If one looks at the Legendre family $y^2=x(x-1)(x-\lambda)$ over the λ-sphere, it is easy to see that the vanishing cycles α and β coming from the bad fibers at $\lambda=0$

and $\lambda=1$ form a basis for $\pi_1(X_t)$. Since $\bar{X}$ is the pull-back of this family via τ, α (resp. β) is a vanishing cycle for $\bar{X}$ at any fiber X_s where $\tau(s)=0$ (resp. $\tau(s)=1$). Since the vanishing cycles of $\bar{X}$ actually vanish in $\pi_1(\bar{X})$, we see that $\pi_1(X_t)\to\pi_1(\bar{X})$ is zero. $\square$

We next turn to homology. We will always work with $\mathbf{Z}$ coefficients, and our first result gives sufficient conditions for $\bar{f}_*$ to be an isomorphism on H_1:

(1.40) **Proposition.** *Let $\bar{f}\colon \bar{X}\to\bar{S}$ be an elliptic fibration with non-trivial local monodromy, and let m_i, $i=1,\dots,\ell$, be the multiplicities of the multiple singular fibers of $\bar{f}$. If $\ell\geqq 2$, assume that the m_i are pairwise relatively prime. Then $\bar{f}_*\colon H_1(\bar{X})\to H_1(\bar{S})$ is an isomorphism.*

This is actually an immediate consequence of the more general proposition (1.41) below.

Given any collection of integers $m_i>1$, $i=1,\dots,\ell$, let $G=G(m_1,\dots,m_\ell)$ be the cokernel of the map $\mathbf{Z}\to\bigoplus_{i=1}^{\ell}\mathbf{Z}/m_i\mathbf{Z}$ (if $\ell=0$, set $G=0$). Note that $G=0$ if and only if $\ell\leqq 1$, or $\ell\geqq 2$ and for $i\neq j$, m_i and m_j are relatively prime (this is the Chinese Remainder Theorem).

The following seems to be well-known (see, for example, [6]).

(1.41) **Proposition.** *Assume that $\bar{f}$ has non-trivial local monodromy and let m_i, $i=1,\dots,\ell$, be the multiplicities of the multiple singular fibers of $\bar{f}$. Then we have an exact sequence:*

$$0\to G(m_1,\dots,m_\ell)\to H_1(\bar{X})\to H_1(\bar{S})\to 0.$$

Proof. This time we take the homology of (1.1):

$$\begin{tikzcd}
 & & 0 & 0 & \\
0 \arrow[r] & M_3 \arrow[r] & H_1(\bar{X}) \arrow[r] \arrow[u] & H_1(\bar{S}) \arrow[r] \arrow[u] & 0 \\
 & H_1(X_t) \arrow[r] \arrow[u] & H_1(X) \arrow[r] \arrow[u] & H_1(S) \arrow[r] \arrow[u] & 0 \\
 & & M_2 \arrow[r] \arrow[u] & M_1 \arrow[u] & \\
 & & 0 \arrow[u] & 0 \arrow[u] &
\end{tikzcd} \tag{1.42}$$

The M_i are the appropriate kernels, and the diagram has the same exactness properties as (1.38) (the second row of (1.42) is exact by the Serre spectral sequence for $f\colon X\to S$). The map $H_1(X_t)\to M_3$ is zero by (1.39), so that we get an exact sequence:

$$M_2\to M_1\to H_1(\bar{X})\to H_1(\bar{S})\to 0.$$

M_1 is the abelian group generated by g_s, $s \in \Sigma$ (each g_s is a loop about s) subject to the single relation $\sum_{s \in \Sigma} g_s = 0$. The proof of (1.37) shows that the image of M_2 in M_1 contains the elements $m_i g_{s_i}$, $i = 1, \ldots, \ell$ and g_s, $s \notin \{s_1, \ldots, s_\ell\}$ (this is the notation of (1.37)). These elements generate a subgroup of M_1 whose quotient is $G = G(m_1, \ldots, m_\ell)$. Thus we get an exact sequence:

(1.43) $G \to H_1(\bar{X}) \to H_1(\bar{S}) \to 0.$

Using the usual presentation of $\pi_1(S)$ with generators α_i, β_i and g_s, $s \in \Sigma$, we get a surjective homomorphism $\pi_1(S) \to G$ by sending g_{s_i} to the image of the generator of $\mathbf{Z}/m_i\mathbf{Z}$ in G, and sending all other generators to zero. This gives us a ramified covering $\bar{S}' \to \bar{S}$ with group G, unramified outside $\{s_1, \ldots, s_\ell\}$, and the ramification index at points above s_i divides m_i. If we construct $\bar{X}'$ as in the proof of (1.36), we see that $\bar{X}' \to \bar{X}$ is a covering space with group G. This is classified by a map $\pi_1(\bar{X}) \to G$, which gives a map $H_1(\bar{X}) \to G$ since G is abelian. The map $G \to H_1(\bar{X})$ from (1.43), followed by this map, is the identity on G, so that $G \to H_1(\bar{X})$ is injective. $\square$

Here is a partial converse to (1.40):

(1.44) **Proposition.** *If $\bar{f}_*$: $H_1(\bar{X}) \to H_1(\bar{S})$ is an isomorphism, then the m_i are pairwise relatively prime (see* (1.40)) *and the monodromy is non-trivial.*

Proof. The proof of (1.41) shows that even if $H^1(X_t) \to H^1(\bar{X})$ is non-zero, we still have an inclusion $G \to \operatorname{Ker}(H_1(\bar{X}) \to H_1(\bar{S}))$. Thus $G = 0$ and the m_i are pairwise relatively prime. If the monodromy is trivial, then $b_1 \geqq 2g + 1$ by Theorem 14.7 of [8]. But $b_1 = 2g$ when $\bar{f}_*$ is an isomorphism. $\square$

Note that (1.44) says nothing about *local* monodromy. This is because having non-trivial local monodromy is not a necessary condition for $\bar{f}_*$ to be an isomorphism, yet just non-trivial monodromy is not enough (i.e., the converse to (1.44) is false). To see this, consider the following examples:

(1.45) *Examples.* Let $\bar{S}$ be an elliptic curve, so that $\pi_1(\bar{S}) \simeq \mathbf{Z} \oplus \mathbf{Z}$. We will construct two elliptic surfaces over $\bar{S}$.

1. Fix a period τ_0 and let j be the constant function $j(\tau_0)$ on $\bar{S}$. Define ρ: $\pi_1(\bar{S}) \to SL(2, \mathbf{Z})$ by $\rho((1,0)) = \rho((0,1)) = -I$. Then ρ gives a locally constant sheaf G on $\bar{S}$ which belongs to j (see [8, §8]). Let $\bar{X}$ be the basic member of $\mathscr{F}(j, G)$. Then $\bar{f}$: $\bar{X} \to \bar{S}$ is smooth (so that $\bar{f}$ has trivial local monodromy), and the Serre spectral sequence gives us an exact sequence:

(1.46) $H_2(\bar{X}) \to H_2(\bar{S}) \to H_1(X_t)_{\pi_1(\bar{S})} \to H_1(\bar{X}) \to H_2(\bar{S}) \to 0.$

Since $H_1(X_t)_{\pi_1(\bar{S})} \simeq \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}$ (an easy computation) and $H_2(\bar{X}) \to H_2(\bar{S})$ is onto ($\bar{f}$ has a section), we see that $\bar{f}_*$: $H_1(\bar{X}) \to H_1(\bar{S})$ is not an isomorphism.

2. Fix the period $\tau_0 = (1 + i\sqrt{3})/2$ and let j be the constant $j(\tau_0) = 0$. Define ρ: $\pi_1(S) \to SL(2, \mathbf{Z})$ by:

$$\rho((1,0)) = \rho((0,1)) = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}.$$

This gives a locally constant sheaf G on $\bar{S}$ that belongs to j [8, § 8], and again we take $\bar{X}$ to be the basic member of $\mathscr{F}(j, G)$. $\bar{f}: \bar{X} \to \bar{S}$ is smooth and there is no local monodromy, but this time one computes that $H_1(X_t)_{\pi_1(\bar{S})} = 0$. Thus, by (1.46), $\bar{f}_*: H_1(\bar{X}) \to H_1(\bar{S})$ is an isomorphism.

These examples lead to another partial converse to (1.40):

(1.47) **Proposition.** *Suppose that $\bar{f}: \bar{X} \to \bar{S}$ has no multiple singular fibers and that $j \neq 0, 1$. Then the following are equivalent:*

1. *$\bar{f}_*: H_1(\bar{X}) \to H_1(\bar{S})$ is an isomorphism.*
2. *$\bar{f}$ has non-trivial local monodromy.*

Proof. $2 \Rightarrow 1$ follows from (1.40). To prove $1 \Rightarrow 2$, assume that $\bar{f}$ is minimal and has trivial local monodromy. Then $\bar{f}$ is smooth because there are no multiple singular fibers, and j is a constant. Thus, in the monodromy representation $\rho: \pi_1(\bar{S}) \to SL(2, \mathbf{Z})$, every $\rho(\gamma)$ has a fixed point τ_0 (the period) on $\mathfrak{h}$. Since $j \neq 0, 1$, $\rho(\gamma)$ is $\pm I$, so that $\rho(\gamma) - I$ is either 0 or $-2I$. Then $H_1(X_t)_{\pi_1(\bar{S})} \simeq \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}$ because the monodromy is non-trivial by (1.44). Since $H_2(\bar{S}) \simeq \mathbf{Z}$, we cannot have a surjection $H_2(\bar{S}) \to H_1(X_t)_{\pi_1(\bar{S})}$. Because $\bar{f}$ is smooth we have the exact sequence (1.46), and this shows that $\bar{f}_*: H_1(\bar{X}) \to H_1(\bar{S})$ is not an isomorphism. □

Here is a useful corollary of all the above:

(1.48) **Corollary.** *Let $\bar{f}: \bar{X} \to \bar{S}$ be an elliptic fibration with non-trivial local monodromy. The following are equivalent:*

1. *$H_1(\bar{X})$ is torsion-free.*
2. *The Néron-Severi group $NS(\bar{X})$ is torsion-free.*
3. *All of the integral homology and cohomology groups $H_i(\bar{X})$ and $H^i(\bar{X})$ are torsion-free.*
4. *$\bar{f}_*: H_1(\bar{X}) \to H_1(\bar{S})$ is an isomorphism.*
5. *If $\bar{f}$ has $\geqq 2$ multiple singular fibers, then their multiplicities are pairwise relative prime.*

Proof. $4 \Leftrightarrow 5$ follows from (1.40) and (1.44), and $1 \Leftrightarrow 2$ is well known. $3 \Rightarrow 1$ is trivial and $1 \Rightarrow 3$ is an easy application of Poincaré duality and the universal coefficient theorem. $1 \Rightarrow 4$ follows from (1.41) (G is finite) while $4 \Rightarrow 1$ is true because $H_1(\bar{S})$ is torsion-free. □

Results similar to these have been obtained independently by R. Mandelbaum [12]. Also see [11, 7] and [9] for a deeper look at the topology of elliptic surfaces.

§ 2. Examples

Before we given the examples, we need to recall the arithmetic aspects of our situation. In § 1 we had a minimal elliptic fibration $\bar{f}: \bar{X} \to \bar{S}$ which has a section σ_0 and a non-constant j-invariant. The generic fiber of $\bar{f}$, a smooth elliptic curve over $K(\bar{S})$ (the function field of $\bar{S}$), can be defined by a cubic equation in $\mathbf{P}^2_{K(\bar{S})}$:

(2.1) $y^2 z = 4x^3 - g_2 x z^2 - g_3 z^3, \quad g_2, g_3 \in K(\bar{S})$.

We will think of this as a point at infinity, (0, 1, 0), together with an affine part defined by the Weierstrass equation:

(2.2) $y^2 = 4x^3 - g_2 x - g_3, \quad g_2, g_3 \in K(\bar{S})$.

Because $\bar{f}$ is minimal, it is the Néron model of (2.2). This means that $\bar{f}: \bar{X} \to \bar{S}$ (up to a fiber preserving isomorphism over $\bar{S}$) and the Weierstrass equation (2.2) (up to an isomorphism $(x, y) \to (u^2 x, u^3 y)$, $u \in K(\bar{S})^*$, which transforms (2.2) into the equation:

(2.3) $y^2 = 4x^3 - g_2' x - g_3', \quad g_2' = u^4 g_2, \quad g_3' = u^6 g_3$)

mutually determine each other.

Furthermore, the group of $K(\bar{S})$-rational solutions of (2.2) (with the point at infinity as identity) is naturally isomorphic to the group $\mathfrak{S}$ of sections of $\bar{f}: \bar{X} \to \bar{S}$ (with σ_0 as identity). We will often speak of solutions and Weierstrass equations rather than sections and elliptic fibrations.

A. The examples we give are all applications of the following:

(2.4) **Theorem.** *Given the following data:*

1. *A Weierstrass equation* (2.2) *over* $\mathbf{C}(t)$ *with* $p_g = 0$,
2. *Solutions* $\sigma_1, \ldots, \sigma_r$ *of* (2.2),
3. *The order of* $\mathfrak{S}_{\mathrm{tor}}$,

there is an effective algorithm (described below) to decide whether or not the σ_i *are a basis of* $\mathfrak{S}$ *modulo torsion.*

Working over $\mathbf{C}(t)$ means that $\bar{S} = \mathbf{P}^1$, and the reason for requiring $p_g = 0$ will soon become clear. Before giving the algorithm, let's note that p_g is easy to compute: from (12.6) and (12.7) of [8] we have:

$$(2.5) \quad \sigma_0^2 = -(p_g - q + 1) = -(1/12)[\deg j + 6 \textstyle\sum \nu(I_b^*) + 2\nu(\mathrm{II}) + 10\nu(\mathrm{II}^*) + 3\nu(\mathrm{III}) + 9\nu(\mathrm{III}^*) + 4\nu(\mathrm{IV}) + 8\nu(\mathrm{IV}^*)]$$

where $\nu(\mathrm{I}_b^*)$, $\nu(\mathrm{II})$, etc. are the numbers of bad fibers of types I_b^*, II, etc. (and these numbers are easy to determine from g_2 and g_3 – see [14, III.17]). Since $q = 0$ by (1.48), we can find p_g.

Now we give the algorithm. Since $p_g = 0$, $\mathfrak{S}$ has rank $-4 + 2(\#\Sigma) - (\sum_{b>0} \nu(\mathrm{I}_b))$ by (3.23). This must equal the number r of given solutions; otherwise they can't form a basis. When we do have the right number of solutions, then by (1.26), the σ_i generate modulo torsion if and only if:

$$\det \langle \sigma_i, \sigma_j \rangle = (\# \mathfrak{S}_{\mathrm{tor}})^2 / \prod_{s \in \Sigma} m_s$$

(m_s is defined in (1.22)).

Thus, we need an effective method to compute $\langle\sigma, \sigma'\rangle$ for σ and σ' in $\mathfrak{S}$. By (1.18), this means computing $(\sigma-\sigma_0)\cdot(\sigma'-\sigma_0)$ (intersection product on $\bar{X}$) and $\sigma\cdot D_s(\sigma')$ for $s\in\Sigma$. To compute the latter, we only have to determine which components of X_s get hit by σ and σ' (see (1.19)). In 2*B* below we give an effective method for doing this. The computation of $(\sigma-\sigma_0)\cdot(\sigma'-\sigma_0)$ is discussed in 2*C* below.

Using this algorithm to compute the examples in 2*E* is quite straightforward. The only tricky part is determining $\#\mathfrak{S}_{\mathrm{tor}}$, and we discuss this in 2*D* below.

B. The problem of determining which component of X_s, $s\in\Sigma$, is hit by an element $\sigma\in\mathfrak{S}$ is evidently local on $\bar{S}$. Thus we can assume that we have a solution $\sigma=(\alpha,\beta)$, $\alpha,\beta\in\mathbf{C}((t))$, of a Weierstrass equation:

(2.6) $y^2=4x^3-g_2x-g_3,\qquad g_2, g_3\in\mathbf{C}((t)).$

This equation has a local Néron model $\bar{f}\colon\bar{X}\to\operatorname{Spec}(\mathbf{C}[[t]])$. We assume that the special fiber X_s of $\bar{f}$ is not smooth, and we say that the Weierstrass equation (2.6) has type I_b, I_b^*, etc. if X_s has that type.

The solution σ gives a section of $\bar{X}$ over $\operatorname{Spec}(C[[t]])$, and we want to know which component of X_s it hits. The difficulty is that constructing $\bar{X}$ from (2.6) is non-trivial (see [14]). However, a first approximation to $\bar{X}$ is fairly easy to obtain. Since we are only interested in (2.6) up to an isomorphism as described in (2.3), we can transform (2.6) into a Weierstrass equation with the following property:

(2.7) *Definition.* A Weierstrass equation (2.6) is called *minimal* if $\operatorname{ord} g_2\geqq 0$, $\operatorname{ord} g_3\geqq 0$ and $\operatorname{ord}\Delta=\operatorname{ord}(g_2^3-27g_3^2)$ is as small as possible (i.e., given an isomorphic equation (2.3) with $\operatorname{ord} g_2'\geqq 0$, $\operatorname{ord} g_3'\geqq 0$, then $\operatorname{ord}\Delta'\geqq\operatorname{ord}\Delta$).

Take $\bar{X}$ and collapse all of the non-zero components to a point (their intersection matrix is negative definite, so this is possible). This gives a local surface $\bar{Y}$ over $\operatorname{Spec}(\mathbf{C}[[t]])$, and:

(2.8) **Lemma.** *The local surface $\bar{Y}$ is defined by any minimal Weierstrass equation (made projective).*

Proof. Néron [14, III.16] shows that $\bar{Y}$ is defined by any standard equation (see [14, III.7]). A minimal equation is standard except in cases I_b and I_b^* $(b>0)$, but in (2.18) and (2.24) below, we show that minimal equations of these types are isomorphic, over $\operatorname{Spec}(\mathbf{C}[[t]])$, to standard ones. □

Thus Y_s (the special fiber of $\bar{Y}$) is the cubic $y^2=4x^3-g_2(0)x-g_3(0)$, which has a unique singular point $(a,0)$ (and $a=0$ except in case I_b). Our solution $\sigma=(\alpha,\beta)$ gives sections of $\bar{X}$ and $\bar{Y}$ over $\operatorname{Spec}(\mathbf{C}[[t]])$ which are compatible with the collapsing map $\pi\colon\bar{X}\to\bar{Y}$. Determining where σ hits Y_s is very easy: just evaluate (α,β) at $t=0$. From this one easily proves:

(2.9.) **Proposition.** *A solution (α,β) of a minimal Weierstrass equation hits a non-zero component of X_s if and only if* $\operatorname{ord}(\alpha-a)>0$. □

We next describe which non-zero component of X_s get hits. We do this case by case, in the following order (based on increasing difficulty): III, III*, IV, IV*,

I_0^*, I_b $(b>0)$ and I_b^* $(b>0)$. Types II and II* are omitted because they have only one component of multiplicity one.

When we write an equation like:

$$\alpha = c t^k + \dots$$

we mean that the omitted terms have degree $>k$.

For a minimal equation of type III or III*, we are done by (2.9): for these types, X_s has precisely one non-zero component C_1 of multiplicity one (see (1.13)).

Next, we consider minimal Weierstrass equations of types IV and IV*. Using [14, III.17], one easily sees that these equations can be written:

$$(2.10)\quad \begin{aligned} &\mathrm{IV}: \ y^2 = 4x^3 - r t^2 x - s t^2 \\ &\mathrm{IV}^*: \ y^2 = 4x^3 - r t^3 - s t^4 \end{aligned}$$

where $r, s \in \mathbf{C}[[t]]$ and $s(0) \neq 0$.

(2.11) **Lemma.** *If* (α, β) *is a solution of* (2.10) *with* $\operatorname{ord}\alpha > 0$, *then:*

$$\beta^2 = \begin{cases} -s(0)t^2 + \dots & \text{Type IV} \\ -s(0)t^4 + \dots & \text{Type IV}^*. \end{cases}$$

Proof. For type IV*, write $\alpha = u t^k$, where u is a unit and $k \geqq 1$. Then we get:

$$\beta^2 = 4u^3 t^{3k} - r u t^{3+k} - s t^4.$$

If $k=1$, this becomes $\beta^2 = (a \text{ unit}) \cdot t^3$, which is impossible. Thus $k>1$ and β^2 is as desired. The argument for type IV is similar and even easier. □

(2.12) **Proposition.** *Suppose we have a minimal Weierstrass equation of type IV or IV*, as in* (2.10). *Pick a square root q of* $-s(0)$. *Then the non-zero components C_1 and C_2 of multiplicity one in X_s (see* (1.13)) *can be labeled so that a solution* (α, β) *of* (2.10) *hits* C_1 (resp. C_2) *if and only if:*

1. (Type IV) $\beta = q t + \dots$ (resp. $\beta = -q t + \dots$)
2. (Type IV*) $\beta = q t^2 + \dots$ (resp. $\beta = -q t^2 + \dots$).

Proof. We will treat type IV – the proof for type IV* is similar. A solution of (2.10) misses the zero component if and only if $\operatorname{ord}\alpha > 0$, and by (2.11), we then have $\beta = \pm q t + \dots$. Thus, we must show that two solutions (α, β) and (α', β') hit the same non-zero component if and only if β and β' have the same coefficient of t. The crucial fact is that the components of multiplicity one form a group (isomorphic to $\mathbf{Z}/3\mathbf{Z}$), and this group structure is compatible with the addition of sections. Thus (α, β) and (α', β') hit the same component if and only if $(\alpha, \beta) - (\alpha', \beta')$ hits the zero component. Set $(\alpha_1, \beta_1) = (\alpha, \beta) - (\alpha', \beta')$, and recall that:

$$(2.13)\quad \alpha_1 = -\alpha - \alpha' + (1/4)[(\beta + \beta')/(\alpha - \alpha')]^2.$$

First, assume that β and β' have the same coefficient of t (which we can assume to be q). Then $(\beta + \beta')/(\alpha - \alpha') = (2qt + \dots)/(\alpha - \alpha')$, so that $\operatorname{ord}(\beta + \beta')/(\alpha - \alpha') \leqq 0$, which implies $\operatorname{ord}\alpha_1 \leqq 0$. Thus, (α, β) and (α', β') do hit the same component.

If β and β' have different coefficients of t, then β and $-\beta'$ have the same coefficients. Since $-(\alpha', \beta')=(\alpha', -\beta')$ the above paragraph shows that (α, β) and $-(\alpha', \beta')$ hit the same component, which forces (α, β) and (α', β') to hit different ones. □

The next case is a minimal Weierstrass equation of type I_0^*. Using [14, III.17] one sees that it can be written:

(2.14) $y^2=4x^3-rt^2x-st^3$

where r, $s\in\mathbf{C}[[t]]$ and $r(0)^3-27s(0)^2\neq 0$. Then the cubic $4x^3-r(0)x-s(0)$ has three distinct roots which we call r_1, r_2 and r_3.

(2.15) **Proposition.** *Suppose we have a minimal Weierstrass equation of type* I_0^* *as in* (2.14). *Then the non-zero components* C_1, C_2 *and* C_3 *of multiplicity one in* X_s (*see* (1.13)) *can be labeled so that a solution* (α, β) *of* (2.14) *hits* C_i *if and only if* $\alpha = r_i t + \ldots$.

Proof. Let (α, β) be a solution of (2.14) that misses the zero component. Using (2.9) to write $\alpha=ut+\ldots$, we get $\beta^2=(4u^3-r(0)u-s(0))t^3+\ldots$, so that $u\in\{r_1, r_2, r_3\}$ and $\operatorname{ord}\beta\geqq 2$. Then one proceeds as in the proof of (2.12), taking the difference of two solutions and using (2.13). □

We move on to the case of a minimal Weierstrass equation of type I_b, $b>0$, which can be written:

(2.16) $y^2=4x^3-g_2x-g_3$

where g_2, $g_3\in\mathbf{C}[[t]]^*$ and $\operatorname{ord}\Delta=b$ (where $\Delta=g_2^3-27g_3^2$). Y_s is defined by $y^2=4x^3-g_2(0)x-g_3(0)$ and has a singular point $(a, 0)$, where $a=-3g_3(0)/2g_2(0)$.

(2.17) **Lemma.** *Let* (α, β) *be a solution of* (2.16) *that misses the zero component. Then* $\operatorname{ord}(12\alpha^2-g_2)>0$. *If we write* $12\alpha^2-g_2=ct^k+\ldots$, $c\neq 0$, *then:*

1. *If b is odd, then* $2k<b$ *and* $\beta^2=(c^2/48a)t^{2k}+\ldots$
2. *If b is even and* $2k<b$, *then* $\beta^2=(c^2/48a)t^{2k}+\ldots$.

Proof. From (2.9) we know that $\alpha(0)=a$. Since $g_2(0)=12a^2$, we see that $\operatorname{ord}(12\alpha^2-g_2)>0$. Then write $g_2=12v^2$ where $v=a+\ldots$, so that $\Delta=g_2^3-27g_3^2=27(8v^3+g_3)(8v^3-g_3)$. If $\Delta=mt^b+\ldots$, then $8v^3+g_3=(m/27\cdot 16a^3)t^b+\ldots$ (since $g_3(0)=-8a^3$). Also $12\alpha^2-g_2=12(\alpha+v)(\alpha-v)$, so that $\alpha-v=(c/24a)t^k+\ldots$.

Let $f(x)=4x^3-g_2x-g_3$. Then we compute that $f(v)=-(8v^3+g_3)$ and $f'(v)=0$, so that:

$$
\begin{aligned}
(2.18)\quad \beta^2=f(\alpha)&=f(v)+f'(v)((c/24a)t^k+\ldots)\\
&\quad+(1/2)f''(v)((c/24a)t^k+\ldots)^2+\ldots\\
&=-(m/27\cdot 16a^3)t^b+\ldots+(c^2/48a)t^{2k}+\ldots
\end{aligned}
$$

and the lemma follows easily. □

(2.19) **Proposition.** *Suppose we have a minimal Weierstrass equation of type* I_b, $b>0$, *as in* (2.16), *and pick a square root* q *of* $3a$. *Then the non-zero components* $C_1, \ldots, C_{b-1}$ *of* X_s (*see* (1.13)) *can be labeled so that if a solution* (α, β) *of* (2.16) *misses the zero component, then:*

1. (α, β) *hits* $C_{b/2}$ *if and only if* $2\operatorname{ord}\beta \geqq b$.
2. (α, β) *hits* C_k *or* C_{b-k} *if and only if* $\operatorname{ord}\beta = k$, $2k < b$.

In this case, we can write $12\alpha^2 - g_2 = c t^k + \ldots$, $c \neq 0$, *and then* (α, β) *hits* C_k (resp. C_{b-k}) *if and only if* $\beta = (c/4q)\, t^k + \ldots$ (resp. $\beta = -(c/4q)\, t^k + \ldots$).

Proof. This proof will make extensive use of [14], including notation. Let v be as in (2.17). Then the change of coordinates $(x, y, z) \to (x - vz, y, z)$ transforms (2.16) (made projective) into the equation:

$$(2.20) \quad A\colon\ y^2 z = 4x^3 + 12 v x^2 z - (8v^3 + g_3) z^3.$$

Let $N = 12v$, $M = -(8v^3 + g_3)$. Then the Néron model $\bar{X}$ is built from the equations

$$A_i\colon\ y^2 z = 4t^i x^3 + N x^2 z + M t^{-2i} z^3$$

for $1 \leqq i \leqq \ell$, where $b = 2\ell$ or $2\ell + 1$ (see [14, §§9–10]). If A_i^0 is the special fiber of A_i, then A_i^0 is defined by $z(y^2 - 12ax^2) = 0$ for $2i < b$ and $A_{b/2}^0$ is defined by $z(y^2 - 12ax^2 + (m/27 \cdot 16a^3) z^2) = 0$.

We have projection maps $\pi_i\colon X_s \to A_i^0$, and we label the components $C_1, \ldots, C_{b-1}$ so that $\pi_{b/2}$ takes $C_{b/2}$ onto the conic:

$$(2.21) \quad y^2 - 12ax^2 + (m/27 \cdot 16a^3) z^2 = 0$$

and π_i takes C_i (resp. C_{b-i}) onto the line $y = 2qx$ (resp. $y = -2qx$). See [14, §9] and the table of generic points [14, p. 104].

The solution (α, β) of (2.16) gives a solution $(\alpha - v, \beta, 1)$ of (2.20) and hence a solution $\sigma_i = (\alpha - v, \beta, t^i)$ of A_i. We have to figure out where $\sigma_i(0)$ lands on A_i^0.

If $\operatorname{ord}\beta = k$, $2k < \beta$, by (2.17) we can write $12\alpha^2 - g_2 = c t^k + \ldots$, $c \neq 0$, and then the proof of (2.17) shows that

$$\begin{aligned}\sigma_k &= ((c/24a)\, t^k + \ldots,\ \pm(c/4q)\, t^k + \ldots, t^k)\\ &= ((c/24a) + \ldots,\ \pm(c/4q) +, 1).\end{aligned}$$

When $t = 0$, this is on either $y = 2qx$ or $y = -2qx$ in A_k^0, depending on the sign of β.

When $2\operatorname{ord}\beta = k$, the argument is similar: one uses (2.17) and (2.18) to verify that $\sigma_{b/2}$ hits $A_{b/2}^0$ on the conic (2.21). The final case, also treated similarly, is when $2\operatorname{ord}\beta > k$. $\square$

Finally, we have the case of a minimal Weierstrass equation of type I_b^*, $b > 0$, which can be written:

$$(2.22) \quad y^2 = 4x^3 - r t^2 x - s t^3$$

where $r, s \in \mathbf{C}[[t]]^*$ and $\operatorname{ord}(r^3 - 27s^2) = b$. Let $r^3 - 27s^2 = m t^b + \ldots$. The cubic $4x^3 - r(0)x - s(0)$ has a double root $a = -3s(0)/2r(0)$ (and the other root is $-2a$).

(2.23) **Lemma.** *Let* (α, β) *be a solution of* (2.22) *which misses the zero component of* X_s. *Then either* $\alpha = -2at + \ldots$ *or* $\alpha = at + \ldots$, *and if* $\alpha = at + \ldots$, *then:*

1. *When b is odd,* $\beta^2 = (-m/27 \cdot 16a^3)\, t^{b+3} + \ldots$
2. *When b is even,*

$$12\alpha^2 - g_2 = \pm(\sqrt{m}/3a)\, t^{(b+4)/2} + \ldots.$$

Proof. We know that $\operatorname{ord}\alpha > 0$, and then $\alpha = -2at + \ldots$ or $\alpha = at + \ldots$ follows as in (2.15).

Since $r(0) = 12a^2$, we can write $r = 12v^2$ where $v = a + \ldots$. Set $\alpha - vt = ct^k + \ldots$, where $c \neq 0$ and $k > 1$. Then manipulating as in the proof of (2.17), we get:

(2.24) $\beta^2 = (-m/27 \cdot 16a^3)\, t^{b+3} + \ldots + 12ac^2 t^{2k+1} + \ldots.$

From this, the lemma follows easily. □

(2.25) **Proposition.** *Suppose we have a minimal Weierstrass equation of type* I_b^*, $b > 0$, *as in* (2.22). *Let q be a square root of* $-m/3a$ (*b* odd) *or of m* (*b* even). *Then the non-zero components* C_1, C_{b+3}, C_{b+4} *of multiplicity one in* X_s (*see* (1.13)) *can be labeled so that if* (α, β) *is a solution of* (2.22) *missing the zero component, then:*

1. (α, β) *hits* C_1 *if and only if* $\alpha = -2at + \ldots.$
2. (α, β) *hits* C_{b+3} *if and only if* $\alpha = at + \ldots$ *and* $\beta = (q/12a)\, t^{(b+3)/2} + \ldots$ (*b* odd) *or* $12\alpha^2 - g_2 = (q/3a)\, t^{(b+4)/2} + \ldots$ (*b* even).
3. (α, β) *hits* C_{b+4} *if and only if* $\alpha = at + \ldots$ *and* $\beta = -(q/12a)\, t^{(b+3)/2} + \ldots$ (*b* odd) *or* $12\alpha^2 - g_2 = -(q/3a)\, t^{(b+4)/2} + \ldots$ (*b* even).

Proof. Let v be as in (2.23). Then $(x, y, z) \to (x - vtz, y, z)$ transforms (2.22) into:

(2.26) $A\colon\ y^2 z = 4x^3 - 12vtx^2 z - (8v^3t^3 + g_3)\, z^3.$

Then $\bar{X}$ is built out of equations $A_1, \ldots, A_{b+2}$, each of which is a transform of A (see [14, III.12]). The cases b even and b odd are treated separately. In each case, the proof of this proposition is similar to the proof of (2.19).

The component C_1 maps onto a component of A_2^0 under $\pi_2\colon \bar{X} \to A_2$; similarly C_{b+3} and C_{b+4} correspond to components of A_{b+2}^0 under π_{b+2} (see [13, III.12]). A solution σ of (2.22) gives solution σ_2 of A_2 and σ_{b+2} of A_{b+2}. Using (2.23) and (2.24), it is easy to see which components these hit. □

Remark. In the proofs of (2.12) and (2.15) (types IV, IV* and I_0^*), we used results from Néron [14] on the number of components of multiplicity one and their group structure. Our methods can be used to give elementary (though cumbersome) proofs of these results and results for some other types (for example, the group structure of the four components of multiplicity one in a fiber of type I_b^*). One the other hand, one can use Néron to prove (2.12) and (2.15) (see (2.19) and (2.25)).

It is clear that all of the above results are valid over any algebraically closed field of characteristic different from 2 or 3.

C. Given σ and σ' in $\mathfrak{S}$, we need to compute the intersection number $\sigma \cdot \sigma'$, where we regard σ and σ' as curves on $\bar{X}$. If $\sigma = \sigma'$, then $\sigma \cdot \sigma' = \sigma'^2 = \sigma_0^2$, and this is easy to compute by (2.5). If $\sigma \neq \sigma'$, then $\sigma \cdot \sigma'$ is a sum of local contributions from the points P of $\sigma \cap \sigma'$ on $\bar{X}$.

The difficulty is that, in practice, we don't have $\bar{X}$; rather, we have a Weierstrass equation (2.2) and two explicit solutions σ and σ'. But one can still compute $\sigma \cdot \sigma'$ from this data.

If σ and σ' meet at a point P which lies either on a good fiber or the zero component of a bad fiber, then near P, $\bar{X}$ is defined by any minimal Weierstrass equation. Then one can use this equation to compute the local contribution at P to $\sigma \cdot \sigma'$.

If σ and σ' meet at a point P on a non-zero component of X_s, $s \in \Sigma$, there are two things one can do. First, one could use Neron [14] and push σ and σ' down to the appropriate A_i and compute these (see the proofs of (2.19) and (2.25) for examples of "pushing down"). The other way is to note that $\sigma - \sigma'$ (the difference in $\mathfrak{S}$) meets σ_0 at infinity on the zero-component of X_s, and the local contribution of $(\sigma - \sigma') \cdot \sigma_0$ at infinity is the same as the local contribution of $\sigma \cdot \sigma'$ at P (subtracting σ' fiberwise gives a birational map of $\bar{X}$ to itself defined in a neighborhood of P, so it preserves local intersection multiplicities).

D. We do not know a systematic method for determining torsion. But there are several tactics which work well. The first is the fact that for $s \in \Sigma$ we have (by [16, 19]) an injection:

(2.27) $\quad \mathfrak{S}_{\mathrm{tor}} \to (X_s')_{\mathrm{tor}}.$

If X_s is not of type I_b, then X_s' is an extension of $\mathbf{C}$ by the finite group G_s of components of multiplicity one. So in this case we have an injection:

(2.28) $\quad \mathfrak{S}_{\mathrm{tor}} \to G_s.$

For example, consider the equation $y^2 = 4x^3 - 3t^3 x + t^4$ over $\mathbf{C}(t)$. The bad fibers occur at $t = 0, 1$ and ∞ and are of types IV^*, I_1 and III respectively. The fiber over $t = 0$ (resp. $t = \infty$) tells us, via (2.28), that any torsion has order 3 (resp. 2). Thus $\mathfrak{S}$ has no torsion.

If all the bad fibers are of type I_b, then (2.27) doesn't give much information. But $\mathfrak{S}_0$ is torsion-free (see (1.17)), so that we still have an injection:

$$\mathfrak{S}_{\mathrm{tor}} \to \bigoplus_{s \in \Sigma} G_s.$$

so all torsion is killed by the l.c.m. of the exponents of the G_s.

Finally, the bilinear form $\langle \ , \ \rangle$ of §1 can be used to give more detailed information. For example, consider the equation $y^2 = x(x-1)(x - t^2 - c)$ over $\mathbf{C}(t)$, where $c \in \mathbf{C} - \{0, 1\}$. The bad fibers occur at $t = \pm\sqrt{-c}$, $\pm\sqrt{1-c}$ and ∞ and are of type I_2 except for $t = \infty$ which is of type I_4. Let σ be a torsion solution, and assume $2\sigma \neq \sigma_0$. Then 2σ hits the zero component except at $t = \infty$ where it must hit C_2. Thus $\langle 2\sigma, 2\sigma \rangle = -(2\sigma - \sigma_0)^2 - 1$ (using (1.25) and (1.26)), so $\langle 2\sigma, 2\sigma \rangle = 1 + 2(2\sigma \cdot \sigma_0)$ since $\sigma^2 = \sigma_0^2 = -1$. Since 2σ is torsion, $\langle 2\sigma, 2\sigma \rangle = 0$ and we get a contradiction. Thus, the only torsion solutions are the obvious 2-torsion ones: $(0,0)$, $(1,0)$, $(t^2 + c, 0)$.

E. Now that the algorithm is complete we give the examples. We will do the first two in detail and then just list the others.

In all these examples one has $q=0$ and, via (2.5), $p_g=0$ and $\sigma_0^2=-1$.

Example 1. Consider the equation over $\mathbf{C}(t)$:

$$(2.29)\quad y^2=4x^3-3t(t-B)^2x-t(t-B)^3$$

where $B\in\mathbf{C}-\{0,1\}$. Here $\Delta=27t^2(t-1)(t-B)^6$ and $j=t/(t-1)$. The bad fibers occur at $t=0,1,B$ and ∞, and are respectively of types II, I_1, I_0^* and III. $\mathfrak{S}$ has rank 3 and no torsion by (2.28) (use the fiber of type II). The algorithm (2.4) shows that a basis of $\mathfrak{S}$ will have determinant (with respect to $\langle\ ,\ \rangle$) equal to 1/8.

Let r_1, r_2, and r_3 be the distinct roots of $4x^3-3Bx-B$, and let s_i be a square root of $-4r_i^3/B$. Then $\sigma_i=(r_i(t-B), s_i(t-B)^2)$ is a solution of (2.29). We claim that these form a basis of $\mathfrak{S}$.

Over $t=0$ and $t=1$ it is clear that σ_0, σ_1, σ_2 and σ_3 hit the fiber in different places.

At $t=B$, we use $T=t-B$ as a local parameter, so (2.29) becomes $y^2=4x^3-3T^2(B+T)x-T^3(B+T)$ (which is minimal) and σ_i is (r_iT, s_iT^2). Then (2.15) shows that each σ_i hits a different component over $t=B$.

At $t=\infty$, we transform the equation to $y^2=4x^3-3t(t-B)^{-2}x-t(t-B)^{-3}$, and using $T=(t-B)^{-1}$ as a local parameter, this becomes the minimal equation $y^2=4x^3-3T(1+BT)x-T^2(1+BT)$, with solutions $\sigma_i=(r_iT, s_iT)$. By (2.9), all the σ_i hit C_1. To see if they meet there, set $(\alpha,\beta)=\sigma_i-\sigma_j$. Then $\alpha=-(r_i-r_j)T+1/4(s_i+s_j)^2(r_i-r_j)^{-2}$, which does not have a pole at $T=0$. Thus, the σ_i do not meet over ∞.

It is also clear that σ_0, σ_1, σ_2, σ_3 cannot meet anywhere else, so that $\sigma_i\cdot\sigma_j=0$ for $i\neq j$, $0\leqq i,j\leqq 3$. Then from (1.18) and (1.19) we see that the matrix $\langle\sigma_i,\sigma_j\rangle$ is:

$$\begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$$

which has determinant 1/8, as desired.

Example 2. Consider the equation over $\mathbf{C}(t)$:

$$(2.30)\quad y^2=x(x-1)(x-t^2-c)$$

where $c\in\mathbf{C}-\{0,1\}$. Since $\Delta=16\lambda^2(\lambda-1)^2$ (where $\lambda=t^2+c$) the bad fibers are at $t=\pm\sqrt{-c}$, $\pm\sqrt{1-c}$ and ∞, all of type I_2 except for $t=\infty$, which is of type I_4. The rank is 1, and in 2*D* we determined the torsion. Thus σ is a basis modulo torsion if and only if $\langle\sigma,\sigma\rangle=1/4$.

Let $m=\sqrt{1-c}-\sqrt{-c}$. Then $\sigma=(m(t+\sqrt{-c}),\ im(t+\sqrt{-c})(t-\sqrt{1-c}))$ is a solution of (2.30). We need to show that $\langle\sigma,\sigma\rangle=1/4$.

The coordinate change $(x,y)\to(x-(\lambda+1)/3, 2y)$ transforms (2.28) to the Weierstrass equation:

$$\begin{aligned}(2.31)\quad y^2=4x^3&-(4(\lambda^2-\lambda+1)/3)x\\ &-4(\lambda+1)(\lambda-2)(2\lambda-1)/27\end{aligned}$$

where $\lambda = t^2 + c$. This is minimal at $t = \pm\sqrt{-c}$, $\pm\sqrt{1-c}$. Our solution becomes $\sigma = (m(t+\sqrt{-c}) - (\lambda+1)/3, 2im(t+\sqrt{-c})(t-\sqrt{1-c}))$.

At $t = \pm\sqrt{-c}$, Y_s has singular point $(-1/3, 0)$. Thus, at $t = \sqrt{-c}$, σ hits C_0, while at $t = -\sqrt{-c}$ σ hits C_1 (using (2.9)). At $t = \pm\sqrt{1-c}$, Y_s has a singular point $(1/3, 0)$. Since $m(\pm\sqrt{1-c} + \sqrt{-c}) = 1$ or $-m^2$, σ hits C_1 at $t = \sqrt{1-c}$ while at $t = -\sqrt{1-c}$, it hits C_0 (again, this is (2.9)).

At $t = \infty$, we use $T = 1/t$ as local parameter, and (2.31), after multiplying g_2 (resp. g_3) by T^4 (resp. T^6), becomes $y^2 = 4x^3 - ((4/3) + \ldots)x - ((8/27) + \ldots)$. Y_s has a singular point $(-1/3, 0)$. σ becomes $(-(1/3) + mT + \ldots, 2imT + \ldots)$, so that by (2.19) σ hits C_1.

It is clear that $\sigma \cdot \sigma_0 = 0$, and then we easily get $\langle \sigma, \sigma \rangle = 1/4$. Thus σ generates modulo torsion, and we know what the torsion is.

The table below gives six more examples. In each of theses cases $\mathfrak{S}_{\mathrm{tor}} = 0$, so we actually give a basis for $\mathfrak{S}$.

(2.32) Solutions of Weierstrass equations over $\mathbf{C}(t)$

Equation	*Basis of* $\mathfrak{S}$
1. $y^2 = 4x^3 - 3t^3x + t^4$	$(0, t^2)$
2. $y^2 = 4x^3 - 3tx + t$	$(1/3, \sqrt{4/27})$
3. $y^2 = 4x^3 - 3tx + 1$	$(0, 1)$
4. $y^2 = 4x^3 - 4t^2x + t^2$	$(0, t), (t, t)$
5. $y^2 = 4x^3 - 3t(t-1)^2x$	$(t-1, 2i(t-1)^2)$,
$\quad - t(t-1)^3$	$(-(1/2)(t-1), (1/\sqrt{2})(t-1)^2)$
6. $y^2 = 4x^3 - 4t^2x + 4$	$(0, 2), (t, 2), (-1, 2t)$,
	$(\lambda, 2\lambda^2 t), \lambda = e^{\pi i/3}$.

Determining that the torsion is zero in Eqs. 3 and 6 is similar to what was done for Example 2 above; in the other cases it is a simple consequence of (2.28).

Another example is a generic elliptic surface with $p_g = 0$. Here $\bar{X}$ is $\mathbf{P}^2$ blown up at nine points on a cubic, and there are 12 singular fibers of type I_1. The 9 exceptional curves are sections σ_i, $0 \leqq i \leqq 8$ (where σ_0 is the zero section). Then one computes that $\det \langle \sigma_i, \sigma_j \rangle_{1 \leqq i,j \leqq 8} = 9$, so that the σ_i generate a subgroup of index 3 in $\mathfrak{S}$.

Remark. The solutions of Examples 1 and 2 and Eqs. 1, 2 and 5 of (2.32) were found by W. Hoyt (who conjectured that they were bases). In each of these cases, C. Schwartz [17] showed that the solutions are independent in $\mathfrak{S}$, and that the solution σ of Example 2 does generate modulo torsion [18]. His methods require involved calculations with Picard-Fuchs equations and automorphic forms.

§3. Parabolic Cohomology, Automorphic Forms and Hodge Theory

A. As in §1, we let $f: X \to S$ be the smooth part of an elliptic fibration, with non-constant j-invariant and zero-section σ_0. The universal covering space of S may be identified with the upper half-plane $\mathfrak{h}$. Pulling back X to $\mathfrak{h}$, one obtains a diagram

$$\begin{tikzcd} \tilde{X} \arrow[r] \arrow[d, "g"'] & X \arrow[d, "f"] \\ \mathfrak{h} \arrow[r, "\pi"] & S \end{tikzcd}$$

There is a simple description of $\tilde{X}$ in terms of periods. Let $\mathscr{F}^1 = f_* \Omega^1_{X/S}$ (the first Hodge filtration bundle over S) and let ω_0 be any generating section of $\tilde{\mathscr{F}}^1 = \pi^* \mathscr{F}^1 = g_* \Omega^1_{\tilde{X}/\mathfrak{h}}$. Choose a basis $\{u_1, u_2\}$ for the constant sheaf $R^1 g_* \mathbf{Z}$ with cup-product $(u_1, u_2) = -1$. Then $\omega = (u_1, \omega_0)^{-1} \omega_0$ gives value 1 when paired with u_1, and $\tau = -(u_2, \omega)$ is a holomorphic function on $\mathfrak{h}$ with values in $\mathfrak{h}$. For any $z \in \mathfrak{h}$, the numbers 1 and $\tau(z)$ generate the period lattice for $X_{\pi(z)}$, and one obtains the formula $\omega = \tau u_1 + u_2$ as a section of $\tilde{\mathscr{V}} = \mathcal{O}_{\mathfrak{h}}(R^1 g_* \mathbf{C})$. This also allows one to represent the universal cover of $\tilde{X}$ as $\mathfrak{h} \times \mathbf{C}$ in a natural way.

(3.1) *Remark.* A simple example is the case where X is given by the Legendre equation $y^2 = x(x-1)(x-t)$ $(S = \mathbf{P}^1 - \{0, 1, \infty\})$. Here it can be arranged that $\tau(z) = z$.

The elements u_1 and u_2 may be construed as generators for $\tilde{V} = H^1(\tilde{X}, \mathbf{C})$ under the natural isomorphism

$$\tilde{V} \xrightarrow{\sim} H^0(\mathfrak{h}, R^1 g_* \mathbf{C}).$$

With respect to this basis, local sections of $\tilde{\mathscr{V}}$ will be represented by column vector valued functions. We will also identify sections of $\mathcal{O}_S(V)$, where $V = R^1 f_* \mathbf{C}$, with their pullbacks to $\mathfrak{h}$.

The fundamental group Γ of S acts as deck transformations on $\mathfrak{h}$, hence on $\tilde{X}$, inducing the *monodromy representation:*

$$M\colon \Gamma \to \mathrm{Aut}_{\mathbf{Z}}(\tilde{V}) = SL(2, \mathbf{Z})$$
$$M(\gamma) = (\gamma^{-1})^*.$$

The locally constant sheaf V (also $V_{\mathbf{Z}}$, $V_{\mathbf{Q}}$, etc.) may be reconstituted from the monodromy by identifying (z, v) with $(\gamma z, M(\gamma) v)$ in $\mathfrak{h} \times \tilde{V}$.

There is the usual notion of parabolic subgroups of Γ, of which the conjugacy classes are in one-to-one correspondence with the points of Σ. We can identify the sheaf cohomology group used in §1 with parabolic group cohomology (as defined in [21]):

(3.2) **Lemma.** *Let V be any locally constant sheaf on the non-singular algebraic curve S, with fundamental group $\Gamma \subset SL(2, \mathbf{R})$, $\tilde{V}$ the associated Γ-module, and*

$$j\colon S \to \bar{S}$$

the inclusion of S in its smooth completion. Then there is a commutative diagram with exact horizontal rows:

$$\begin{tikzcd} 0 \arrow[r] & H^1_P(\Gamma, \tilde{V}) \arrow[r] \arrow[d, dashed, "\simeq"] & H^1(\Gamma, \tilde{V}) \arrow[r] \arrow[d, "\simeq"] & \bigoplus\limits_{\Gamma_0} H^1(\Gamma_0, \tilde{V}) \arrow[d, "\simeq"] \\ 0 \arrow[r] & H^1(\bar{S}, j_* V) \arrow[r] & H^1(S, V) \arrow[r] & \bigoplus\limits_{s \in \Sigma} H^1(\Delta^*(s), V) \end{tikzcd}$$

where Γ_0 ranges over conjugacy class representatives of parabolic subgroups of Γ, the top row defines the parabolic cohomology group $H_P^1(\Gamma, \tilde{V})$, and $\Delta^*(s)$ is a small punctured disc about the point s.

Proof. (Cf. [25, §12].) □

B. In §1, we defined $\delta(\sigma)$ for $\sigma \in \mathfrak{S}$, and the pairing $\langle \sigma, \sigma' \rangle$. We will now show that these coincide respectively with the cohomology classes of automorphic forms naturally associated to the sections, and the generalized Eichler bilinear form.

We begin by mentioning the way in which the cohomology groups in Lemma (3.2) are computed using differential forms, leaving the full discussion for Part C. On S, we have the holomorphic deRham complex:

$$\Omega_S^{\bullet}(V):\ \mathcal{O}_S(V) \xrightarrow{d} \Omega_S^1(V)$$

forming a resolution of V. This can be extended to a resolution of $j_* V$ by taking:

(3.3) $\bar{\mathscr{V}} \to d\bar{\mathscr{V}}$,

where $\bar{\mathscr{V}}$ denotes the *canonical extension* [2, p. 95] of $\mathscr{V} = \mathcal{O}_S(V)$. Let $\mathscr{V}(*D)$ denote the sheaf of germs of meromorphic sections of $\bar{\mathscr{V}}$ having arbitrary poles only on the support of the fixed effective divisor D. Then:

(3.4) $\bar{\mathscr{V}}(*D) \to d[\bar{\mathscr{V}}(*D)]$

will be called the complex of *forms of the second kind* with values in $\bar{\mathscr{V}}$ and poles on D. Since the complex (3.4) evidently resolves $j_* V$, we have:

(3.5) **Lemma.** *The inclusion of the complex* (3.3) *in* (3.4) *induces an isomorphism on hypercohomology.* □

Thus, we should regard differentials of the second kind as intrinsically representing cohomology classes on $\bar{S}$ (as opposed to S). However, the presence of poles will cause a technical problem in computing cup products.

Let σ be a section of X. Then σ determines a section $\tilde{\sigma}: \mathfrak{h} \to \tilde{X}$, which may be lifted to a mapping:

$$(1, F):\ \mathfrak{h} \to \mathfrak{h} \times \mathbf{C},$$

well-defined up to periods. F may be considered to be an explicit expression for the *normal function* associated to $\sigma - \sigma_0$, regarding X as its own system of Jacobian varieties. One should regard F as giving the image of the vector $F(z)\,u_1$ under the mapping:

$$\psi:\ \tilde{\mathscr{V}} \to (\tilde{\mathscr{F}}^1)^*$$

$$\psi \begin{bmatrix} x \\ y \end{bmatrix} \omega = x - y\tau$$

induced by cup-product. If $\gamma \in \Gamma$ and:

$$M(\gamma) = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

set $j(M(\gamma),\tau)=c\tau+d$. Then it is easy to show that F satisfies the functional equation:

(3.6) $j(M(\gamma^{-1}),\tau)F(\gamma^{-1}z)=F(z)+q_\gamma\tau+r_\gamma$

for suitable integers q_γ, r_γ. Defining the periods for F as:

$$\beta(\gamma)=\begin{bmatrix} r_\gamma \\ -q_\gamma \end{bmatrix}\in\tilde{V},$$

the vector whose image in $(\tilde{\mathscr{F}}^1)^*$ is $q_\gamma\tau+r_\gamma$, one sees immediately using (3.6), that $\beta(\gamma)$ is a 1-cocycle for Γ acting on $\tilde{V}$, so it determines an element of $H^1(\Gamma,\tilde{V})$ (which will soon be seen to be parabolic).

One associates to every section a generalized automorphic form in the sense of Hoyt [4] according to the following recipe. For any function F on $\mathfrak{h}$, we may define its derivative with respect to τ:

$$\frac{dF}{d\tau}=F'(z)\,\tau'(z)^{-1}.$$

Note that $d/d\tau$ is a meromorphic differential operator with poles at the ramification points of τ.

(3.7) **Proposition** [4]. *If F satisfies Eq.* (3.6), *then $G=d^2F/d\tau^2$ is a meromorphic automorphic form of weight* 3 *with respect to* (τ,M), *i.e.*,

$$G(\gamma z)=j(M(\gamma),\tau)^3\,G(z)\qquad\forall\gamma\in\Gamma.\ \square$$

The space of all meromorphic functions satisfying the above transformation rule will be denoted $A_3(M,\tau)$. Similarly, as:

$$\varphi_G=G(z)\begin{bmatrix}\tau\\1\end{bmatrix}d\tau=d\left(dF/d\tau\begin{bmatrix}\tau\\1\end{bmatrix}-F\begin{bmatrix}1\\0\end{bmatrix}\right)$$

is invariant under Γ, it descends to S as a meromorphic 1-form of the second kind with values in V; in fact, (see (3.24)):

$$\varphi_G\in\Gamma(\bar{S},d[\bar{\mathscr{V}}(*D)]),$$

where D is the ramification divisor on $\bar{S}$ for τ. Implicit in this last statement is the assertion that there is a proper notion of ramification at the points of Σ, a matter which will be settled in (3.17). Note also that $\operatorname{ord}_z d\tau=\operatorname{ord}_{\gamma z} d\tau$ for all $\gamma\in\Gamma$, so for $s\in S$, $\operatorname{ord}_s d\tau$ can be defined.

The remainder of §3B will be devoted to the relation among F, φ_G, and $\delta(\sigma)$ [defined in (1.6)]. Let $G\in A_3(M,\tau)$, with φ_G of the second kind, and fix a base point $z_0\in\mathfrak{h}$. The $\tilde{V}$-valued meromorphic function:

(3.8) $\phi(z)=\int_{z_0}^{z}\varphi_G$

is well-defined, independent of choice of path from z_0 to z. Its *period cocycle*, given by:

$$\alpha(\gamma)=M(\gamma)\,\phi(z)-\phi(\gamma z)\in C^1(\Gamma,\tilde{V})$$

represents the deRham cohomology class of φ_G in $H^1(\Gamma, \tilde{V}) \simeq H^1(S, V)$, which we write as $\alpha = [\varphi_G]$. Of course, Lemma (3.5) implies that φ_G naturally represents an element of the parabolic cohomology $H^1(\bar{S}, j_* V)$. If G is a cusp form, this observation is visible at the group cocycle level.

The periods for F and φ_G are related by:

(3.9) **Proposition.** *$\beta = \alpha$ in $H^1(S, V)$.*

Proof. It can be seen directly that we may choose F and ϕ so that:

$$\begin{aligned} F(z) &= {}^t\begin{bmatrix} \tau \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \phi(z) \\ &= (\omega \cup \phi)(z) \qquad \text{(cup-product in } \tilde{V}). \end{aligned}$$

Putting $\Theta = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, we have from (3.6):

$$\begin{aligned} q_\gamma \tau + r_\gamma &= {}^t\omega(z)\, \Theta \phi(z) - j(M\gamma^{-1}, \tau)\, {}^t\omega(\gamma^{-1} z)\, \Theta\, \phi(\gamma^{-1} z) \\ &= {}^t\omega(z)\, \Theta\, \phi(z) - j(M\gamma^{-1}, \tau) [j(M\gamma^{-1}, \tau)^{-1}\, {}^t\omega(z)\, {}^tM\gamma^{-1}] \\ &\quad \cdot \Theta [M\gamma^{-1} \phi(z) - \alpha(\gamma^{-1})] \\ &= {}^t\omega(z)\, {}^tM\gamma^{-1}\, \Theta\, \alpha(\gamma^{-1}) = - {}^t\begin{bmatrix} \tau \\ 1 \end{bmatrix} \Theta\, \alpha(\gamma) \end{aligned}$$

since ${}^tM \Theta M = \Theta$ (flatness of cup-product) and $M\gamma\alpha(\gamma^{-1}) + \alpha(\gamma) = 0$ (cocycle condition). Writing:

$$\alpha(\gamma) = \begin{bmatrix} a_\gamma \\ b_\gamma \end{bmatrix},$$

we obtain

$$q_\gamma \tau + r_\gamma = -(b_\gamma \tau - a_\gamma),$$

or $\beta(\gamma) = \alpha(\gamma)$, as desired. □

Using this we obtain:

(3.10) **Proposition.** *Let σ be a section of X, represented by the holomorphic function F on $\mathfrak{h}$, and put as usual $G = d^2 F / d\tau^2$. Then:*

$$\delta(\sigma) = [\varphi_G]$$

in $H^1(S, V)$.

Proof. In view of (3.9), we need only check that $\delta(\sigma) = \beta$. Let $\{U_k\}$ be an open covering of S consisting of coordinate discs such that $U_j \cap U_k$ is connected. For each j choose a connected component $\tilde{U}_j$ of $\pi^{-1}(U_j)$. Then if $U_j \cap U_k \neq \emptyset$, there is a unique $\gamma_{jk} \in \Gamma$ with:

$$\tilde{U}_j \cap \gamma_{jk} \tilde{U}_k \neq \emptyset.$$

The group cocycle β is sent to the Cech cochain:

$$\hat{\beta}_{jk}=(\pi|_{\tilde{U}_j}^*)^{-1}\,\beta(\gamma_{jk})|_{U_j\cap U_k}.$$

The class $\delta(\sigma)$ is given as a connecting homomorphism in Cech cohomology (cf. [24, § 3]). One lifts σ to $F|_{\tilde{U}_j}$ on U_j, representing an element of $\Gamma(U_j,(\mathscr{F}^1)^*)$. Then $\delta(\sigma)$ is given on $U_j\cap U_k$ by the difference, with $\gamma=\gamma_{jk}$,

$$\begin{aligned}&\operatorname{image}\{F(\gamma^{-1}z)M\gamma u_1-F(z)u_1\}\quad\text{in }(\mathscr{F}^1)^*\\&\quad=F(\gamma^{-1}z)j(M\gamma^{-1},\tau)-F(z)\\&\quad=\operatorname{image}\beta(\gamma)=\hat{\beta}_{jk}.\end{aligned}$$

Thus, $\delta(\sigma)=\beta=[\varphi_G]$ by (3.9). □

Let now $\varphi_1,\ \varphi_2\in\Gamma(\bar{S},d\bar{\mathscr{V}}(*D))$, where – very important – we assume D is supported on S, and let $\phi_1,\ \phi_2$ be their respective integrals (3.8). We will make use of a fundamental domain $\mathscr{D}$ in $\mathfrak{h}$ for the action of Γ, of the type defined in [21]. Its most important feature is that its boundary consists of an even number of smooth arcs (edges) with the property that for any edge $\mathscr{E}$ of the oriented boundary $\partial\mathscr{D}$, there exists a unique $\gamma\in\Gamma$ so that $-\gamma(\mathscr{E})$ is also an edge. Thus we have:

(3.11) **Lemma.** *Let ψ be a Γ-invariant 1-form on $\mathfrak{h}$, integrable on $\partial\mathscr{D}$. Then:*

$$\int_{\partial\mathscr{D}}\psi=0.\quad\square$$

Given φ_1 and φ_2 as above, the *Eichler bilinear form* [4] is defined as:

$$I(\varphi_1,\varphi_2)=\int_{\partial\mathscr{D}}{}^t\phi_1\wedge\Theta\,\varphi_2,$$

where we arrange by choice of $\mathscr{D}$ that $\varphi_1,\ \varphi_2$ are regular on $\partial\mathscr{D}$. If $\varphi_i=\varphi_{G_i}$ with $G_i=d^2F_i/d\tau^2$, then:

$$I(\varphi_1,\varphi_2)=\int_{\partial\mathscr{D}}F_1G_2\,d\tau=\int_{\partial\mathscr{D}}F_2G_1\,d\tau,$$

generalizing the definition given in [3].

(3.12) **Proposition.** $I(\varphi_1,\varphi_2)=[\varphi_1]\cup[\varphi_2]$, *where the cup product:*

$$H^1(\bar{S},j_*V)\times H^1(\bar{S},j_*V)\to H^2(\bar{S},\mathbf{C}).$$

is induced from the cup product on V.

Proof. Both ϕ_j and φ_j are regular in $\bar{\mathscr{V}}$ at the cusps. To remove their poles on D, one makes a correction near $|D|$, obtaining:

$$\eta_j=\varphi_j-d\mu_j,$$

where η_j is now a C^∞ 1-form in $\bar{\mathscr{V}}$. Then:

$$
\begin{aligned}
[\varphi_1] \cup [\varphi_2] &= \int_S {}^t\eta_1 \wedge \Theta\eta_2 \\
&= \int_{\mathscr{D}} {}^t\eta_1 \wedge \Theta\eta_2 \qquad \text{(abusing notation)} \\
&= \int_{\mathscr{D}} {}^t d\lambda_1 \wedge \Theta\eta_2 \qquad \text{write} \quad \eta_j = d\lambda_j \ \text{on}\ \mathfrak{h} \\
&= \int_{\partial\mathscr{D}} {}^t\lambda_1 \wedge \Theta\eta_2 \\
&= \int_{\partial\mathscr{D}} {}^t\phi_1 \wedge \Theta\eta_2 - \int_{\partial\mathscr{D}} (\mu_1 + v_1) \wedge \Theta\eta_2 \\
&\qquad\qquad\qquad\qquad \phi_1 - \lambda_1 + \mu_1 = v_1, \qquad \text{a constant} \\
&= \int_{\partial\mathscr{D}} {}^t\phi_1 \wedge \Theta\eta_2 \qquad \text{(Stokes' theorem and (3.11))} \\
&= \int_{\partial\mathscr{D}} {}^t\phi_1 \wedge \Theta\eta_2 - \int_{\partial\mathscr{D}} {}^t\phi_1 \wedge \Theta(d\mu_2) \\
&= I(\varphi_1, \varphi_2) - \int_{\partial\mathscr{D}} d({}^t\phi_1 \wedge \Theta\mu_2) + \int_{\partial\mathscr{D}} \varphi_1 \wedge \Theta\mu_2 \\
&= I(\varphi_1, \varphi_2);
\end{aligned}
$$

the above integrals all converge because the integrands have at worst logarithmic singularities at the cusps, as follows from the properties of $\mathscr{V}$. □

Putting (3.10) and (3.12) together, we have shown that the Eichler pairing of automorphic forms coming from sections of an elliptic surface – as studied by Hoyt and Schwartz ([4, 5, 18]) – are computable in an elementary manner using intersection numbers, as described in §§ 1, 2. It deserves to be repeated that in (3.12) we (and they) assume that τ is unramified at the cusps, lest the Eichler pairing be undefined.

C. We will develop the Hodge theory necessary to give a unified treatment of several known or conjectured results concerning parabolic cohomology and automorphic forms. We begin by quoting the relevant facts from [25], always taking V to be $R^1 f_* \mathbf{C}$ for an elliptic surface.

To begin, we know that the complexes (3.3) and (3.4) resolve $j_* V$. The Hodge theory for $H^i = H^i(\bar{S}, j_* V)$ can be described by a decreasing filtration F on either of the complexes (call it $K^{\cdot}$):

$$K^{\cdot} = F^0 \supset F^1 \supset F^2 \supset F^3 = 0,$$

with successive quotients $\mathrm{Gr}_F^p = F^p/F^{p+1}$, so that the induced filtration $\{F^p H^i\}$ on cohomology gives a Hodge structure of weight $i+1$. Deligne calls this kind of data a *cohomological Hodge complex* of weight one; specifically, it requires:

(3.13) (i) The spectral sequence

$$E_1^{p,q} = \mathbf{H}^{p+q}(\bar{S}, \mathrm{Gr}_F^p K^{\cdot}) \Rightarrow \mathbf{H}^{p+q}(\bar{S}, K^{\cdot}) = H^{p+q}$$

degenerates at E_1. Then

(a) $\mathbf{H}^i(\bar{S}, F^p K^{\cdot})$ maps injectively into H^i (with image $F^p H^i$, of course);

(b) There is a canonical identification of $\mathrm{Gr}_F^p H^i$ with $\mathbf{H}^i(\bar{S}, \mathrm{Gr}_F^p K^{\cdot})$.

(ii) $H^{p,q}=F^pH^i\cap\overline{F^qH^i}$ projects isomorphically onto $\mathrm{Gr}_F^p H^i$ for $p+q=i+1$.

The main result of [25] gives the existence of a filtration on the complex (3.3), making it a cohomological Hodge complex for $j_* V$, such that it yields the same Hodge structure on H^i as the one induced by the classical Hodge theory on $\bar{X}$. Let $\bar{\mathscr{F}}^1$ be the sub-bundle of $\bar{\mathscr{V}}$ determined by $\mathscr{F}^1$. Denote by Σ' the subset of Σ consisting of points where the monodromy is not unipotent; thus $s\in\Sigma'$ if and only if X_s is singular and not of type I_b. The filtration on (3.3) is (see [25, §9]):

$$\begin{aligned}(3.14)\quad & F^2\colon\ 0\to\Omega_{\bar S}^1(\log\Sigma')\otimes\bar{\mathscr{F}}^1\\ & F^1\colon\ \bar{\mathscr{F}}^1\to d\bar{\mathscr{V}}\\ & F^0\colon\ \bar{\mathscr{V}}\to d\bar{\mathscr{V}}\end{aligned}$$

hence we have:

$$\begin{aligned}(3.15)\quad & \mathrm{Gr}_F^1\colon\ \bar{\mathscr{F}}^1\to\Omega_{\bar S}^1(\log\Sigma)\otimes\overline{\mathscr{G}r}^0\\ & \mathrm{Gr}_F^0\colon\ \overline{\mathscr{G}r}^0\to 0\end{aligned}$$

where $\overline{\mathscr{G}r}^0$ denotes the quotient $\bar{\mathscr{V}}/\bar{\mathscr{F}}^1$.

Before proceeding further, we need to discuss the ramification of τ on $\bar{S}$. At a point $s\in S$, where the fiber is smooth, we have the usual notion; in terms of a local parameter t centered at s,

$$\tau=\tau(0)+u\,t^m,\quad \text{where}\quad u(0)\neq 0,$$

which we write as:

$$\begin{aligned}&\tau=\tau(0)+(t^m);\\ &d\tau=(t^{m-1})\,dt.\end{aligned}$$

For the singular fibers, there are the following formulae [8, §8] in a suitably chosen local parameter:

Fiber type	τ		$d\tau$	
I_0^*	$\tau_0+(t^m)$		$(t^{m-1})\,dt$	
I_b, I_b^* $(b>0)$	$\dfrac{1}{2\pi i}\log t$		$(1)\,dt/t$	(3.16)
II, IV*	$\eta+(t^{h/3})$	$\eta=e^{2\pi i/3}$ $h\equiv 1(3)$	$(t^{h/3})\,dt/t$	
II*, IV	$\eta+(t^{h/3})$	$h\equiv 2(3)$	$(t^{h/3})\,dt/t$	
III, III*	$i+(t^{h/2})$	$h\equiv 1(2)$	$(t^{h/2})\,dt/t$	

As a first guess, ramification should indicate excessive vanishing for $d\tau$. However, we will soon see that we must alter this preconception slightly in order to obtain the proper notion. What we do is to look at the size of the cokernel of d in the complex $\mathrm{Gr}_F^1(\bar{\mathscr{V}}\to d\bar{\mathscr{V}})$ (3.15). Computing it is a relatively easy matter, for in Gr_F^1 d becomes $\mathcal{O}_S$-linear, so it suffices to differentiate a generator $\bar{\omega}$ of $\bar{\mathscr{F}}^1$. At a regular fiber, $\bar{\mathscr{F}}^1$ is generated by $\omega=\begin{bmatrix}\tau\\1\end{bmatrix}$, so

$$d(\omega)=(t^{m-1})\,dt\begin{bmatrix}1\\0\end{bmatrix},$$

so $\mathrm{ord}_s\, d(\omega)$ and $\mathrm{ord}_s(d\tau)$ coincide.

(3.17) *Definition.* If $s\in S$, the *order of ramification* of τ at s is:

$$\mu=\mu(s)=m=1+\dim[(\Omega_S^1\otimes\overline{\mathscr{G}\imath}^0)/d\bar{\mathscr{F}}^1]_s;$$

if $s\in\Sigma$, then set:

$$\mu=\dim[(\Omega_S^1(\log\Sigma)\otimes\overline{\mathscr{G}\imath}^0)/d\bar{\mathscr{F}}^1]_s,$$

where we are introducing the notation:

$$[\mathscr{S}]_s=\mathscr{S}\otimes\mathcal{O}_s$$

for $\mathcal{O}_S$-modules $\mathscr{S}$.

Note that it is possible to compute $\mu(s)$ by working at any pre-image of s in $\mathfrak{h}$. In order to calculate the ramification orders at points of Σ, we must determine a generator $\bar{\omega}$ for each of the possible fiber types. The following chart can be verified, using the same notation as before:

	Type	Monodromy	Eigenvalue/ Eigenvector pairs	$\bar{\omega}$	
1.	I_0^*	$-\begin{pmatrix}1&0\\0&1\end{pmatrix}$		$t^{1/2}\begin{bmatrix}\tau\\1\end{bmatrix}$	
2.	I_b	$\begin{pmatrix}1&b\\0&1\end{pmatrix}$		$\begin{bmatrix}\tau\\1\end{bmatrix}$	
3.	I_b^* $(b>0)$	$-\begin{pmatrix}1&b\\0&1\end{pmatrix}$		$t^{1/2}\begin{bmatrix}\tau\\1\end{bmatrix}$	
4.	II	$\begin{pmatrix}1&1\\-1&0\end{pmatrix}$	$-\eta,\ A=\begin{bmatrix}\eta\\1\end{bmatrix};\ -\bar{\eta},\ B=\begin{bmatrix}\bar{\eta}\\1\end{bmatrix}$	$t^{5/6}[A-t^{h/3}B]$	(3.18)
	II*	$\begin{pmatrix}0&-1\\1&1\end{pmatrix}$	$-\eta,\ A;\ -\eta,\ B$	$t^{1/6}[A-t^{h/3}B]$	
	IV	$\begin{pmatrix}0&1\\-1&-1\end{pmatrix}$	$\bar{\eta},\ A;\ \eta,\ B$	$t^{2/3}[A-t^{h/3}B]$	
	IV*	$\begin{pmatrix}-1&-1\\1&0\end{pmatrix}$	$\eta,\ A;\ \bar{\eta},\ B$	$t^{1/3}[A-t^{h/3}B]$	
	III	$\begin{pmatrix}0&1\\-1&0\end{pmatrix}$	$-i,\ \hat{A}=\begin{bmatrix}i\\1\end{bmatrix};\ i,\ \hat{B}=\begin{bmatrix}-i\\1\end{bmatrix}$	$t^{3/4}[\hat{A}-t^{h/2}\hat{B}]$	
	III*	$\begin{pmatrix}0&-1\\1&0\end{pmatrix}$	$i,\ \hat{A};\ -i,\ \hat{B}$	$t^{1/4}[\hat{A}-t^{h/2}\hat{B}]$	

Thus, we obtain for the four sub-divisions above:

(3.19) 1. $d(\bar{\omega})=t^{1/2}\begin{bmatrix}1\\0\end{bmatrix}d\tau=t^m\begin{bmatrix}1\\0\end{bmatrix}dt/t, \quad \mu=m>0$

2. $d(\bar{\omega})=\begin{bmatrix}1\\0\end{bmatrix}dt/t, \quad \mu=0$

3. $d(\bar{\omega})\equiv t^{1/2}\begin{bmatrix}1\\0\end{bmatrix}dt/t, \quad (\mathrm{mod}\,\bar{\mathscr{F}}^1), \quad \mu=0$

4. In each of the six types, we have

$$\bar{\omega}=t^{\alpha}[v-t^{\beta}\bar{v}]$$

for suitable α, β, and v. Then

$$\begin{aligned} d(\bar{\omega}) &\equiv -(\alpha+\beta)\,t^{\alpha+\beta}\,\bar{v}\,dt/t \\ &= t^{2\alpha+\beta-1}(t^{1-\alpha}\bar{v})\,dt/t. \end{aligned}$$

Thus, $\mu=2\alpha+\beta-1$. If β takes on the minimum possible value (1/3, 2/3, or 1/2), then μ is either 0 or 1, depending on whether $\alpha<1/2$ or $\alpha>1/2$ respectively ("star" or "non-star" fiber types).

We now begin to describe the Hodge structure on $H^1=H^1(\bar{S},j_*V)$.

(3.20) **Proposition**

(1) $F^2H^1\simeq\{G\in A_3(M,\tau)$: *$\varphi_G$ is a holomorphic section of* $\Omega^1_{\bar{S}}(\log\Sigma')\otimes\bar{\mathscr{F}}^1\}$.

(2) $h^{1,1}=\dim H^{1,1}=\sum_{s\in\Sigma}\mu(s)+\sum_{s\in S}[\mu(s)-1]$.

Proof. As $F^2H^1\simeq H^1(\bar{S},F^2K^{\cdot})\simeq H^0(\bar{S},\Omega^1_{\bar{S}}(\log\Sigma')\otimes\bar{\mathscr{F}}^1)$, (1) follows immediately. For (2), we have:

$$H^{1,1}\simeq\mathbf{H}^1(\bar{S},\mathrm{Gr}^1_F K^{\cdot}).$$

Let C be the cokernel of d in $\mathrm{Gr}^1_F K^{\cdot}$. Writing $\mathrm{Gr}^1_F K^{\cdot}$ as $A\to B$, we have a short exact sequence of complexes:

$$0\to(A\to\operatorname{im}d)\to(A\to B)\to(0\to C)\to 0.$$

As the first term is acyclic, we obtain:

$$H^{1,1}\simeq\mathbf{H}^1(\bar{S},(0\to C))\simeq H^0(\bar{S},C),$$

and thus:

$$h^{1,1}=\dim H^0(\bar{S},C)=\sum_{s\in\Sigma}\mu(s)+\sum_{s\in S}[\mu(s)-1]. \quad \square$$

As an immediate consequence of (3.20), we obtain a conceptually clear proof of a result in [23] attributed to Kodaira:

(3.21) **Corollary**

$$h^{1,1} \geqq \nu(\mathrm{I}_0^*) + \nu(\mathrm{II}) + \nu(\mathrm{III}) + \nu(\mathrm{IV}).$$

Proof. One sees from (3.19) that at each $s \in \Sigma$, where a fiber of type I_0^*, II, III, or IV occurs, $\mu(s) \geqq 1$. □

If $\sigma \in \mathfrak{S}$, its image under δ is a class of type $(1,1)$ in $H^1(\bar{S}, j_* V)$. This follows immediately from the fact that $\delta(\sigma)$ is the reduction of the fundamental class of an algebraic cycle (1.6) – which is of type $(1,1)$ in $H^2(\bar{X})$ – since the Hodge structure is compatible with the Leray spectral sequence. When $p_g(\bar{X}) = 0$, then $H^1(\bar{S}, j_* V)$ is purely of type $(1,1)$; by the Hodge conjecture for surfaces (Lefschetz existence theorem), it is generated by the image of δ, so (3.21) gives a lower bound for the rank of $\mathfrak{S}$.

When $\bar{X}$ is an elliptic modular surface, i.e., if $\tau(z) = z$, we obtain examples of the Hodge structures of Shimura (see [25, §12]). As it is described in [22], the projection:

$$S_3(\Gamma) \to H_P^1(\Gamma, \tilde{V}) \xrightarrow{\mathrm{Re}} H_P^1(\Gamma, \tilde{V}_{\mathbf{R}})$$

is an isomorphism; here, the space of cusp forms $S_3(\Gamma)$ gives $H^{2,0}$, and it is clear that $H^{1,1} = 0$.

We can also give a direct discussion of the following formula given by Shioda:

(3.22) **Proposition** [23]. *Let*

r = rank of $\mathfrak{S}$

g = genus of $\bar{S}$

ν = number of singular fibers

ν_1 = number of fibers of type I_b

p_g = geometric genus of $\bar{X}$.

Then $r \leqq 4g - 4 + 2\nu - \nu_1 - 2p_g$.

Proof. Since the homomorphism δ embeds $\mathfrak{S}/\mathfrak{S}_{\mathrm{tor}}$ in $H^{1,1}$, it suffices to show that the right-hand side of the inequality is equal to $h^{1,1}$. Now:

$$\begin{aligned} h^{1,1} &= \dim H^{1,1} = \dim H^1 - 2\dim H^{2,0} \\ &= \dim H^1 - 2p_g \end{aligned}$$

because the Hodge structure is compatible with the Leray spectral sequence, and those on $H^0(\bar{S}, R^2 \bar{f}_* \mathbf{C})$ and $H^2(\bar{S}, R^0 \bar{f}_* \mathbf{C})$ are purely of type $(1,1)$. Then using the exact sequence:

$$0 \to H^1 \to H^1(S, V) \to H^0(\bar{S}, R^1 j_* V) \to H^2 = 0$$

and the facts:

$$H^0(\bar{S}, R^1 j_* V) \simeq \bigoplus_{s\in\Sigma} (V/N_s V),$$

where

$$N_s = M(\gamma_s) - I$$

$$\gamma_s = \text{parabolic element associated to } s\in\Sigma,$$

$$H^0 = H^0(S, V) = 0 = H^2(S, V),$$

we obtain:

$$\begin{aligned}\dim H^{1,1} &= \dim H^1(S, V) - \sum_{s\in\Sigma} \dim(V/N_s V) - 2p_g \\ &= -\chi(S)\dim \tilde{V} - v_1 - 2p_g \\ &= -2[\chi(\bar{S}) - v] - v_1 - 2p_g \\ &= 4g - 4 + 2v - v_1 - 2p_g. \quad \square\end{aligned}$$

(3.23) *Remark.* When $p_g = 0$, equality holds, for then $r = h^{1,1}$, as was remarked earlier.

The remainder of the paper will be devoted to the verification of the following result:

(3.24) **Theorem.** *Let $\{s_k\}$ be the points of $\bar{S}$ where τ is ramified. Put:*

$$D = \sum_k \mu(s_k) \cdot [s_k].$$

Then, using differential forms of the second kind, $F^1 H^1$ is isomorphic to:

$$\{\varphi \in H^0(\bar{S}, \Omega^1_{\bar{S}}(\log \Sigma') \otimes \bar{\mathscr{F}}^1(D)) : \varphi \text{ is of the second kind}\}.$$

[*Such φ are of the form φ_G, where $G \in A_3(M, \tau)$ may have a pole of order $2\mu(s_k) - 1$ at $s_k \in S$. Also G satisfies the cusp condition at the points of Σ with singular fibers of types* I_b *or* I_b^* $(b > 0)$ (*see* (3.19)).]

W. Hoyt conjectured some cases of this theorem, on the grounds that the dimensions were equal.

Since $F^1 H^1$ contains the image of δ, we would expect, because of (3.10), that if φ comes from a section σ of the elliptic surface, then it lies in subspace appearing in (3.24). This is easy to check. We may assume that the section passes through the identity components of the singular fibers. Since the construction of φ involves taking two τ-derivatives of the function F associated to σ (see §3A), we may freely alter F by periods. Since σ can be expressed as a local single-valued holomorphic function in terms of a generator of $\bar{\mathscr{F}}^{1*}$, where the periods become multiplied by $t^{\alpha-1}$ for some $0 < \alpha \leqq 1$, in all cases F may be written in the form:

$$F = t^{1-\alpha} H(t) \qquad 0 < \alpha \leqq 1,$$

where $H(t)$ is holomorphic.

Considering the issue case by case, one gets:

Type	F	τ	$dF/d\tau$	$\varphi=d(dF/d\tau)\otimes\begin{bmatrix}\tau\\1\end{bmatrix}$
Non-singular	(1)	$\tau_0+(t^m)$	(t^{1-m})	$(t^{-m})\begin{bmatrix}\tau\\1\end{bmatrix}dt$
I_0^*	$(t^{1/2})$	$\tau_0+(t^m)$	$(t^{1/2-m})$	$(t^{-m})\,t^{1/2}\begin{bmatrix}\tau\\1\end{bmatrix}dt/t$
I_b	(1)	$\dfrac{b}{2\pi i}\log t$	(t)	$(1)\begin{bmatrix}\tau\\1\end{bmatrix}dt$
I_b^* $(b>0)$	$(t^{1/2})$	$\dfrac{b}{2\pi i}\log t$	$(t^{1/2})$	$(t^{1/2})\begin{bmatrix}\tau\\1\end{bmatrix}dt/t$
Other	$(t^{1-\alpha})$	$\tau_0+(t^\beta)$	$(t^{1-\alpha-\beta})$	$(t^{1-2\alpha-\beta})\,t^\alpha\begin{bmatrix}\tau\\1\end{bmatrix}dt/t$

The assertion now follows by comparison with our previous calculation of μ (3.19).

The proof of (3.24) depends on putting a Hodge filtration on the complex (3.4) compatible with the one given in (3.14). In other terms we will show that there is a *filtered quasi-isomorphism* between the two complexes. Let K_1 and K_2 be filtered complexes of sheaves with filtrations denoted by F. Then a morphism:

$$\iota\colon K_1^\bullet \to K_2^\bullet$$

is called a filtered quasi-isomorphism if for all p:

(3.25) $\mathrm{Gr}_F^p\iota\colon \mathrm{Gr}_F^p K_1^\bullet \to \mathrm{Gr}_F^p K_2^\bullet$

induces an isomorphism on cohomology sheaves. In the present context, $K_1^\bullet$ will be a subcomplex of $K_2^\bullet$ with the induced filtration, in which case it suffices to show that, for all p, the quotient

$$\mathrm{Gr}_F^p K_2^\bullet/\mathrm{Gr}_F^p K_1^\bullet$$

is acyclic. A filtered quasi-isomorphism induces an isomorphism on hypercohomology and on all filtration levels thereof.

To motivate the Hodge filtration on (3.4), we will introduce the Hodge filtration for the *mixed* Hodge theory for $\bar{S}-|D|$:

(3.26) **Proposition** [25, §13]. *$H^*(\bar{S}-|D|, j_* V)$ is computable as the hypercohomology of the complex:*

$$L^\bullet\colon \bar{\mathscr{V}} \to d\bar{\mathscr{V}}(\log D).$$

Moreover, the filtration:

$$F^2:\ 0\to\Omega_S^1(\log D)\otimes\bar{\mathscr{F}}^1$$

$$F^1:\ \bar{\mathscr{F}}^1\to d\bar{\mathscr{V}}(\log D)$$

$$F^0:\ \bar{\mathscr{V}}\to d\bar{\mathscr{V}}(\log D)$$

induces the Hodge filtration of the natural mixed Hodge structure on the above cohomology groups. □

We first discuss the case where τ is unramified at the cusps. The complexes to be defined differ only on $|D|$ from $L^\cdot$ above. Since the definition of the filtrations are determined locally, as well as the quasi-isomorphism of filtered complexes, we may assume without loss of generality that D consists of one point s where τ is ramified to order m, and $\bar{S}$ is a disc centered at $t=0$.

We first introduce the usual order-of-pole filtration (cf. [2, p. 80]) on:

$$P^\cdot=\Omega_S^\cdot(*D)\otimes\mathscr{V}$$

(induced by that of $\Omega_X^\cdot(*f^{-1}(D))$):

$$F^2:\ 0\to\Omega_S^1(1)\otimes\mathscr{F}^1,$$

$$F^1:\ \mathscr{F}^1\to(\Omega_S^1(1)\otimes\mathscr{V})+(\Omega_S^1(2)\otimes\mathscr{F}^1)$$

for $k\geqq 0$

$$F^{-k}:\ \mathscr{V}(k+1)+\mathscr{F}^1(k+2)\to(\Omega_S^1(k+2)\otimes\mathscr{V})+(\Omega_S^1(k+3)\otimes\mathscr{F}^1)$$

where for a locally-free sheaf $\mathscr{S}$, $\mathscr{S}(n)$ means that one allows poles of order n on D. Actually, the infinite length of the filtration can be circumvented, for the inclusion of $F^0P^\cdot$ in $P^\cdot$ is a quasi-isomorphism, as will follow. We compute:

$$\mathrm{Gr}_F^1 L^\cdot:\ \mathscr{F}^1\to\Omega_S^1(1)\otimes\mathscr{G}\imath^0$$

$$\mathrm{Gr}_F^0 L^\cdot:\ \mathscr{G}\imath^0\to 0$$

$$\mathrm{Gr}_F^1 P^\cdot:\ \mathscr{F}^1(1)\to(\Omega_S^1(1)\otimes\mathscr{G}\imath^0)\oplus[t^{-2}\,dt\,\mathscr{F}^1]_0$$

$$\mathrm{Gr}_F^0 P^\cdot:\ \mathscr{G}\imath^0(1)\oplus[t^{-2}\mathscr{F}^1]_0\to[t^{-2}\,dt\,\mathscr{G}\imath^0]_0\oplus[t^{-3}\,dt\,\mathscr{F}^1]_0$$

for $k>0$

$$\mathrm{Gr}_F^{-k} P^\cdot:\ [t^{-k-1}\mathscr{G}\imath^0]_0\oplus[t^{-k-2}\mathscr{F}^1]_0$$
$$\to[t^{-k-2}\,dt\,\mathscr{G}\imath^0]_0\oplus[t^{-k-3}\,dt\,\mathscr{F}^1]_0.$$

That $(L^\cdot,F)\subset(P^\cdot,F)$ is a filtered quasi-isomorphism follows from the elementary fact that $d(t^{-n})=-nt^{-n-1}\,dt$ which translates cohomologically into the assertion that:

$$[t^{-n}\mathscr{S}]_0\to[t^{-n-1}\,dt\,\mathscr{S}]_0$$

is acyclic; for instance, there is an exact sequence of complexes (which is not, however, split):

$$\begin{aligned}&0\to([t^{-k-1}\mathcal{G}\imath^0]_0\to[t^{-k-2}\,dt\,\mathcal{G}\imath^0]_0)\\&\quad\to(\mathrm{Gr}_F^{-k}P^\bullet)\to([t^{-k-2}\mathcal{F}^1]_0\to[t^{-k-3}\,dt\,\mathcal{F}^1]_0)\to 0\end{aligned}$$

so $\mathrm{Gr}_F^{-k}P^\bullet$ is acyclic for $k>0$.

The above is quite general, for it does not make use of the fact that D is the ramification divisor for τ. We have $\mathcal{F}^1$ generated by the vector $\omega=\begin{bmatrix}\tau\\1\end{bmatrix}$, and:

$$d(\omega)=\begin{bmatrix}1\\0\end{bmatrix}d\tau=\begin{bmatrix}1\\0\end{bmatrix}u\,t^{m-1}\,dt$$

for some invertible power series $u(t)$. Thus, we may improve upon the order of pole designations. Define:

$$\begin{aligned}&\tilde{F}^2\colon\ 0\to\Omega_S^1(1)\otimes\mathcal{F}^1,\\&\tilde{F}^1\colon\ \mathcal{F}^1(m-1)\to(\Omega_S^1(m)\otimes\mathcal{F}^1)+(\Omega_S^1(1)\otimes\mathcal{V}),\\&\tilde{F}^0\colon\ \mathcal{F}^1(m)+\mathcal{V}(1)\to(\Omega^1(m+1)\otimes\mathcal{F}^1)+(\Omega_S^1(2)\otimes\mathcal{V}).\end{aligned}$$

(3.27) **Proposition.** *$(P^\bullet,F)\subset(P^\bullet,\tilde{F})$ is a filtered quasi-isomorphism.*

Proof. We have:

$$\begin{aligned}&\mathrm{Gr}_{\tilde F}^1P^\bullet\colon\ \mathcal{F}^1(m-1)\to([\Omega_S^1(m)/\Omega_S^1(1)]\otimes\mathcal{F}^1)\oplus(\Omega_S^1(1)\otimes\mathcal{G}\imath^0)\\&\mathrm{Gr}_{\tilde F}^0P^\bullet\colon\ [t^{-m}\mathcal{F}^1]_0\oplus\mathcal{G}\imath^0(1)\to[t^{-m-1}\,dt\,\mathcal{F}^1]_0\oplus[t^{-2}\,dt\,\mathcal{G}\imath^0]_0\end{aligned}$$

for $k>0$

$$\mathrm{Gr}_{\tilde F}^{-k}\colon\ [t^{-m-k}\mathcal{F}^1]_0\oplus[t^{-k-1}\mathcal{G}\imath^0]_0\to[t^{-m-k-1}\,dt\,\mathcal{F}^1]_0\oplus[t^{-k-2}\,dt\,\mathcal{G}\imath^0]_0.$$

Comparing $\mathrm{Gr}_{\tilde F}^pP^\bullet$ with $\mathrm{Gr}_F^pP^\bullet$, the result follows by elementary considerations.

It remains now to push the notion of "second kind" (sk) through the filtered complexes we have just defined. We obtain in this manner:

$$\begin{aligned}&(L^\bullet_{\mathrm{sk}},F)=(\Omega^\bullet_S(V),F)\\&(P^\bullet_{\mathrm{sk}},F)\\&(P^\bullet_{\mathrm{sk}},\tilde F).\end{aligned}$$

By almost exactly the same arguments as were used in proving (3.27), we may deduce:

(3.28) **Proposition.** *The inclusions*

$$(\Omega^\bullet_S(V),F)\subset(P^\bullet_{\mathrm{sk}},F)\subset(P^\bullet_{\mathrm{sk}},\tilde F)$$

are filtered quasi-isomorphisms. □

Using (3.28), we can now prove (3.24). We have:

$$\mathrm{Gr}_F^1 P_{\mathrm{sk}}^{\bullet}\colon\ \mathscr{F}^1(m-1)\to([\Omega_S^1(m)_{\mathrm{sk}}/\Omega_S^1]\otimes\mathscr{F}^1)\oplus(\Omega_S^1\otimes\mathscr{G}\imath^0),$$

which we abbreviate as:

$$G^{\bullet}\colon\ D\to A/B\oplus C.$$

Since $d(\omega)=\begin{bmatrix}1\\0\end{bmatrix}d\tau$, $d_2\colon D\to C$ is an isomorphism. If we set:

$$G_1^{\bullet}\colon\ 0\to A/B$$

$$G_2^{\bullet}\colon\ D\xrightarrow{d_2}C,$$

then we have a short exact sequence of complexes:

$$0\to G_1^{\bullet}\to G^{\bullet}\to G_2^{\bullet}\to 0.$$

We should remember at this point that although the definitions of the various complexes is seemingly made on S, they carry, by agreement, a prescribed extension to $\bar{S}$ as in (3.3). By construction, $G_2^{\bullet}$ is acyclic, so:

$$\mathbf{H}^*(\bar{S},G_1^{\bullet})\xrightarrow{\sim}\mathbf{H}^*(\bar{S},G^{\bullet}).$$

Therefore:

$$\mathrm{Gr}_F^1 H^1\simeq\mathbf{H}^1(\bar{S},G^{\bullet})\simeq H^0(\bar{S},A/B).$$

Finally, it follows from the exact sequence:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & H^0(\bar{S},B) & \longrightarrow & H^0(\bar{S},A) & \longrightarrow & H^0(\bar{S},A/B) & \longrightarrow & H^1(\bar{S},B)\\
 & & \wr\| & & & & \wr\| & & \wr\|\\
 & & F^2H^1 & & & & \mathrm{Gr}_F^1 H^1 & \xrightarrow{\ 0\ } & F^2H^2
\end{array}$$

that $F^1H^1\simeq H^0(\bar{S},A)=H^0(\bar{S},\Omega_{\bar{S}}^1(D)_{sk}\otimes\bar{\mathscr{F}}^1)$, which is the desired result (3.24) when $|D|\cap\Sigma=\emptyset$.

The proof is nearly identical when τ has ramification at the cusps. We need only make the appropriate changes in the complex. One obtains for $\tilde{F}^1$ at a point of $\Sigma'\subset\Sigma$ (if a singular fiber is of type I_b, no change from (3.3) is necessary):

$$\tilde{F}^1\colon\ \bar{\mathscr{F}}^1(\mu)\to[\Omega_{\bar{S}}^1(\log\Sigma)]\otimes[\bar{\mathscr{F}}^1(\mu)+\bar{\mathscr{V}}]$$

$$\mathrm{Gr}_F^1\colon\ \bar{\mathscr{F}}^1(\mu)\to[\Omega_{\bar{S}}^1(\log\Sigma)\otimes(\bar{\mathscr{F}}^1(\mu)/\bar{\mathscr{F}}^1)]\oplus[\Omega_{\bar{S}}^1(\log\Sigma)\otimes\overline{\mathscr{G}\imath}^0].$$

The rest of the argument is essentially the same as before; details are left to the reader. □

(3.29) *Remark.* Let us point out how (3.24) relates to the classical description of $F^1H^1(\bar{X},\mathbf{C})$ using meromorphic 2-forms of the second kind. Let w be the variable on $\mathbf{C}$ in the universal cover $\tilde{X}=\mathfrak{h}\times\mathbf{C}$ of X. Then, given $G\in A_3(M,\tau)$

such that $[\varphi_G] \in F^1 H^1(\bar{S}, j_* V)$ (cf. (3.24)), $\psi_G = G(z)\, dw \wedge d\tau$ defines a 2-form on X which is of the second kind on $\bar{X}$. It is easy to see that the cohomology class of ψ_G lies in L^1 and is equal to $[\varphi_G]$ in $L^1/L^2 \simeq H^1(\bar{S}, j_* V)$.

References

1. Clemens, C.H.: Degeneration of Kähler manifolds. Duke Math. J. **44**, 215–290 (1977)
2. Deligne, P.: Equations differentielles à points singuliers réguliers. In: Lecture Notes in Math. 163. Berlin-Heidelberg-New York: Springer 1970
3. Eichler, M.: Eine Verallgemeinerung der Abelschen Integrale. Math. Zeitschr. **67**, 267–298 (1957)
4. Hoyt, W.: Parabolic cohomology and cusp forms of the second kind for extensions of the field of modular functions. Preprint 1978
5. Hoyt, W., Schwartz, C.: Period relations for the Weierstrass equation $y^2 = 4x^3 - 3ux - u$. In preparation (1979)
6. Iitaka, S.: Deformations of compact complex surfaces. In: Global analysis, papers in honor of K. Kodaira. Princeton: Princeton University Press 1969
7. Kas, A.: On the deformation types of regular elliptic surfaces. Preprint 1976
8. Kodaira, K.: On compact analytic surfaces, II–III. Annals of Math. **77**, 563–626; **78**, 1–40 (1963)
9. Kodaira, K.: On homotopy $K-3$ surfaces. In: Essays on topology and related topics, Mémoires dédiés à George deRham. Berlin-Heidelberg-New York: Springer 1970
10. Kodaira, K.: On the structure of compact analytic surfaces, I. Amer. J. Math. **87**, 751–798 (1964)
11. Mandelbaum, R.: On the topology of elliptic surfaces. Preprint 1977
12. Mandelbaum, R.: On the topology of non-simply connected elliptic surfaces with degenerate fibers. Manuscript 1978
13. Manin, Ju.I.: The Tate height of points on an abelian variety; its variants and applications. Amer. Math. Soc. Transl. (Series 2) **59**, 82–110 (1966)
14. Néron, A.: Modèles minimaux des variétés abéliennes sur les corps locaux et globaux. Pub. Math. I.H.E.S. **21** (1964)
15. Néron, A.: Quasi-fonctions et hauteurs sur les variétés abéliennes. Annals of Math. **82**, 249–331 (1965)
16. Ogg, A.: Cohomology of abelian varieties over function fields. Annals of Math. **76**, 185–212 (1962)
17. Schwartz, C.: Independent solutions of certain Weierstrass equations. Manuscript 1977
18. Schwartz, C.: On generators of the group of rational solutions of a certain Weierstrass equation. Trans. A.M.S., to appear
19. Shafarevitch, I.: Principal homogeneous spaces defined over a function field. Amer. Math. Soc. Transl. (Series 2) **37**, 85–114 (1964)
20. Shafarevitch, I., and others: Algebraic Surfaces. Moscow: Proc. Steklov Institute 1965; English Translation, Providence: A.M.S. 1967
21. Shimura, G.: Introduction to the arithmetic theory of automorphic forms. Princeton: Princeton University Press, 1971
22. Shimura, G.: Sur les intégrales attachées aux formes automorphes. J. Math. Soc. Japan **11**, 291–311 (1959)
23. Shioda, T.: On elliptic modular surfaces. J. Math. Soc. Japan **24**, 20–59 (1972)
24. Zucker, S.: Generalized intermediate Jacobians and the theorem on normal functions. Inventiones Math. **33**, 185–222 (1976)
25. Zucker, S.: Hodge theory with degenerating coefficients: L_2 cohomology in the Poincaré metric, Annals of Math. (1979)

Received July 15, 1978/March 26, 1979

Inventiones math. 53, 45–58 (1979)

Inventiones mathematicae

The Structure of Minimizing Hypersurfaces Mod 4

Brian White

Mathematics Department, Fine Hall, Princeton University, NJ 08544, USA

Introduction

The introduction of the generalized surfaces of geometric measure theory has made it possible to prove the existence of solutions to Plateau's problem or, more generally, the existence of surfaces having a given boundary and minimizing an elliptic integral. Naturally that raises the problem of determining the structure of the solutions and, in particular, of proving regularity properties. Of course the nature of the solutions depends on the class of surfaces one allows. The m-dimensional rectifiable flat chains mod ν in $\mathbb{R}^{m+1}$ are characterized by the property that they agree, except on a set of arbitrarily small m-dimensional measure, with the images in $\mathbb{R}^{m+1}$ of C^1 singular m-chains with coefficients in the integers mod ν, the particular chains depending on the degree of approximation desired. Minimizing m-dimensional flat chains mod 2 in $\mathbb{R}^{m+1}$ solve the "unoriented problem" and exhibit the strongest regularity properties; except for closed sets of $m-2$ dimensional measure 0, they are smoothly embedded manifolds [ASS]. Two-dimensional area-minimizing flat chains mod 3 resemble some soap films; in [T] it is shown that they consist of analytic surfaces together with smooth curves along which three such surfaces meet at equal angles.

In this paper we determine the structure of any m-dimensional rectifiable flat chain mod 4 which minimizes the integral of an even elliptic integrand Φ. In particular, we show that, except for a closed set of $m-2$ dimensional measure 0, it is a smoothly immersed manifold, the sheets of which are locally Φ minimizing. In case Φ is analytic, we can also analyze the multiple points of the immersion. For instance, we have as a corollary of the main theorem:

If Q is a two-dimensional area minimizing flat chain mod 4 in $\mathbb{R}^3$, then $\operatorname{spt} Q \sim \operatorname{spt} \partial Q$ is an immersed analytic manifold. Its self-intersection set consists of analytic curves along which two sheets cross transversely, together with isolated points where two sheets are tangent and at which an even number of such curves meet at equal angles (Fig. 1).

0020-9910/79/0053/0045/$02.80

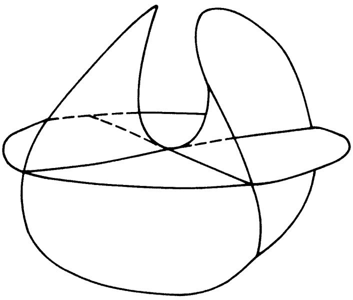


Fig. 1. An area minimizing flat chain mod 4 which decomposes into a disk and enneper's surface

The central idea of the proof is that the mod 4 variational problem is locally equivalent to two mod 2 variational problems. Let Q be an m-dimensional rectifiable flat chain mod 4 in the unit ball $\mathbb{B}^{m+1}$ of $\mathbb{R}^{m+1}$, with boundary ∂Q in the unit sphere $\partial\mathbb{B}^{m+1}$. We show that Q is, as a measure (i.e., ignoring orientations), equal to the sum of two rectifiable flat chains mod 2, Q_1 and Q_{-1}, in $\mathbb{B}^{m+1}$ with boundaries in $\partial\mathbb{B}^{m+1}$. This decomposition is well-behaved with respect to the boundary operator: if Q and Q' are two such flat chains mod 4 and if $\partial Q = \partial Q'$, then $\partial Q_1 = \partial Q'_1$ and $\partial Q_{-1} = \partial Q'_{-1}$. Finally, every pair of such rectifiable flat chains mod 2 arises from the decomposition of a flat chain mod 4. Consequently, Q (in $\mathbb{B}^{m+1}$, with ∂Q in $\partial\mathbb{B}^{m+1}$) is Φ minimizing mod 4 if and only if Q_1 and Q_{-1} are Φ minimizing mod 2. It then follows from the regularity of Φ minimizing flat chains mod 2 that the support of any Φ minimizing flat chain mod 4 is (except for a closed set of $m-2$ dimensional measure 0) locally the union of two smooth manifolds each of which satisfies the Euler-Lagrange equations for Φ.

Figure 2 illustrates a very simple case. The mod 4 boundary, consisting of four semicircles each joining a point to its antipode, is decomposed into two great circles. Minimizing area for the two circles separately gives two disks, which together form the area minimizing flat chain mod 4.

In Fig. 3, the mod 4 boundary is three horizontal circles on a sphere, all oriented from West to East. This decomposes into: (i) the top and bottom circles, and (ii) the middle circle. The corresponding area minimizing flat chains

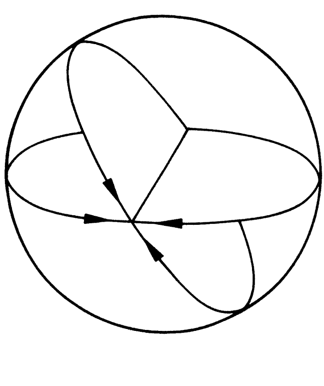


Fig. 2. An area minimizing flat chain mod 4 which decomposes into two disks

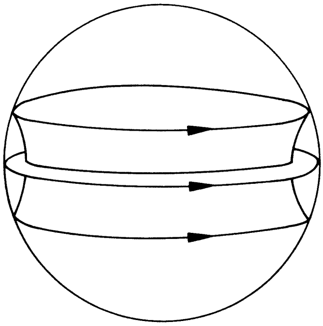


Fig. 3. An area minimizing flat chain mod 4 which decomposes into a catenoid and a disk

mod 2 are a catenoid and a disk. Putting them together gives the area minimizing flat chain mod 4.

I am very grateful to F.J. Almgren, Jr. for his indispensable suggestions and encouragement.
The research was supported by a National Science Foundation Graduate Fellowship.

1. Preliminaries

The outline of this paper is as follows. Chapter 1 contains a list of the standard terminology used. In Chap. 2, Propositions 2.1 and 2.3 prove that every $T \in I_m^\nu(\mathbb{R}^m)$ is the equivalence class mod ν of an integral current. This allows us to apply the very powerful Gauss-Green Theorem of [F1, 4.5.6] to $T \in I_m^\nu(\mathbb{R}^m)$. Propositions 2.4 and 2.6 establish the properties of the mod 4-to-mod 2 decomposition. Finally, Theorems 2.7 and 2.9 apply the decomposition to variational problems mod 4. In general, we use the definitions and notation of [F1]. In particular, we use the following.

1.1. $n \geqq m$ are positive integers.

1.2. If k is an integer and ν is a positive integer, we define

$$|k(\mathrm{mod}\ \nu)| = \min\{|k'|: k' \equiv k\ (\mathrm{mod}\ \nu)\}$$

1.3. For $p \in \mathbb{R}^n$ and $r > 0$, we let

$$\mathbb{B}^n(p,r) = \{x \in \mathbb{R}^n: |x-p| \leqq r\}$$
$$\partial\mathbb{B}^n(p,r) = \{x \in \mathbb{R}^n: |x-p| = r\}$$

1.4. $\mathscr{H}^m$ denotes m-dimensional Hausdorff measure on $\mathbb{R}^n$. If μ is a positive measure on $\mathbb{R}^n$ and $x \in \mathbb{R}^n$ we define the m-dimensional density of μ at x to be

$$\Theta^m(\mu, x) = \lim_{r \downarrow 0} \{\mu\,\mathbb{B}^n(x,r)/(\alpha(m)\,r^m)\}$$

(when this limit exists), where $\alpha(m)$ is the m dimensional volume of the unit ball in $\mathbb{R}^m$.

1.5. $\mathscr{R}_m(\mathbb{R}^n)$ denotes the space of m-dimensional rectifiable currents in $\mathbb{R}^n$. Such currents are characterized by the property that they agree, to within a set of arbitrarily small $\mathscr{H}^m$ measure, with m-dimensional C^1 singular chains with integer coefficients. For $T \in \mathscr{R}_m(\mathbb{R}^n)$, $\|T\|$ denotes the variation measure associated with T [F1, 4.1.5].

Every rectifiable $T \in \mathscr{R}_m(\mathbb{R}^n)$ can be written

$$T = \|T\| \wedge \vec{T}$$

where $\vec{T}$ is a $\|T\|$ measurable m-vector field on $\mathbb{R}^n$ such that $\vec{T}(x)$ is a simple unit m-vector and $\Theta^m((\|T\|, x)$ is a positive integer for $\|T\|$ almost all x.

Also:

$$\|T\| = \mathscr{H}^m \llcorner \Theta^m(\|T\|, \cdot)$$

so that $\|T\|$ is completely determined by its m-dimensional densities [F1, 4.1.28].

1.6. We also define the spaces of integral currents and integral flat chains:

$$I_m(\mathbb{R}^n) = \{T\colon T \in \mathscr{R}_m(\mathbb{R}^n),\ \partial T \in \mathscr{R}_{m-1}(\mathbb{R}^n)\}$$

$$\mathscr{F}_m(\mathbb{R}^n) = \{R + \partial T\colon R \in \mathscr{R}_m(\mathbb{R}^n),\ T \in \mathscr{R}_{m+1}(\mathbb{R}^n)\}.$$

For $T \in \mathscr{F}_m(\mathbb{R}^n)$, $\operatorname{spt} T$ denotes the support of T [F1, 4.1.1], ∂T denotes the boundary of T, and $\mathbf{M}(T)$ denotes the mass of T[F1, 4.1.7]. The *flat norm* $\mathscr{F}(T)$ of T is defined as $\inf\{\mathbf{M}(S) + \mathbf{M}(T - \partial S)\colon S \in \mathscr{R}_{m+1}(\mathbb{R}^n)\}$ [F1, 4.1.24].

$\mathscr{R}_m^{\mathrm{loc}}(\mathbb{R}^n)$ denotes the space of locally rectifiable currents in $\mathbb{R}^n$. $\mathscr{F}_m^{\mathrm{loc}}(\mathbb{R}^n)$ denotes the space of locally integral flat chains in $\mathbb{R}^n$ [F1, 4.1.24].

1.7. Let ν be a nonnegative integer. For $T \in \mathscr{F}_m(\mathbb{R}^n)$, we define:

$$\begin{aligned} \mathscr{F}^\nu(T) &= \inf\{\mathscr{F}(T + \nu Q)\colon Q \in \mathscr{F}_m(\mathbb{R}^n)\} \\ \mathbf{M}^\nu(T) &= \inf\{\mathbf{M}(T + \nu Q)\colon Q \in \mathscr{F}_m(\mathbb{R}^n)\} \\ \operatorname{spt}^\nu(T) &= \bigcap \{\operatorname{spt} R\colon R \in \mathscr{F}_m(\mathbb{R}^n),\ \mathscr{F}^\nu(T - R) = 0\}. \end{aligned}$$

The space of flat chains modulo ν in $\mathbb{R}^n$ is defined by:

$$\mathscr{F}_m^\nu(\mathbb{R}^n) = \mathscr{F}_m(\mathbb{R}^n) / \{T \in \mathscr{F}_m(\mathbb{R}^n)\colon \mathscr{F}^\nu(T) = 0\}.$$

If $T \in \mathscr{F}_m(\mathbb{R}^n)$, we let $T(\operatorname{mod} \nu)$ denote its equivalence class in $\mathscr{F}_m^\nu(\mathbb{R}^n)$.

We also have the space of rectifiable flat chains modulo ν:

$$\mathscr{R}_m^\nu(\mathbb{R}^n) = \mathscr{R}_m(\mathbb{R}^n) / \nu \mathscr{R}_m(\mathbb{R}^n)$$

and the space of rectifiable flat chains modulo ν whose boundaries are rectifiable flat chains modulo ν:

$$I_m^\nu(\mathbb{R}^n) = \{T\colon T \in \mathscr{R}_m^\nu(\mathbb{R}^n),\ \partial T \in \mathscr{R}_{m-1}^\nu(\mathbb{R}^n)\}.$$

For $T\in\mathscr{R}_m^\nu(\mathbb{R}^n)$, $\|T\|^\nu$ denotes the variation measure associated with T [F1, 4.2.26, p. 427]. For $T\in\mathscr{R}_m(\mathbb{R}^n)$, one sets $\|T\|^\nu=\|T(\operatorname{mod}\nu)\|^\nu$. It follows [F1, 4.2.26, p. 430] that for $T\in\mathscr{R}_m(\mathbb{R}^n)$:

1) $\Theta^m(\|T\|^\nu,x)=|\Theta^m(\|T\|,x)(\operatorname{mod}\nu)|$

for $\mathscr{H}^m$ almost all x.

If $T\in\mathscr{R}_m^\nu(\mathbb{R}^n)$, we have:

2) $\|T\|^\nu=\mathscr{H}^m \llcorner \Theta^m(\|T\|^\nu,\cdot)$

This follows from (1.5), and from the existence of a $T'\in\mathscr{R}_m(\mathbb{R}^n)$ such that $\|T\|^\nu=\|T'\|$ and $T=T'+\nu Q$ for some $Q\in\mathscr{R}_m(\mathbb{R}^n)$ [F1, 4.2.26, p. 430].

1.8. Let K be a solid ball or the boundary of a solid ball in $\mathbb{R}^n$. We define:

$$\mathscr{F}_m(K)=\{T\in\mathscr{F}_m(\mathbb{R}^n)\colon \operatorname{spt}T\subset K\}$$

$$\mathscr{R}_m(K)=\{T\in\mathscr{R}_m(\mathbb{R}^n)\colon \operatorname{spt}T\subset K\}$$

$$I_m(K)\ =\{T\in I_m(\mathbb{R}^n)\colon \operatorname{spt}T\subset K\}$$

and similarly

$$\mathscr{R}_m^\nu(K)=\{T\in\mathscr{F}_m^\nu(\mathbb{R}^n)\colon \operatorname{spt}^\nu T\subset K\}$$

$$\mathscr{R}_m^\nu(K)=\{T\in\mathscr{R}_m^\nu(\mathbb{R}^n)\colon \operatorname{spt}^\nu T\subset K\}$$

$$I_m^\nu(K)\ =\{T\in I_m^\nu(\mathbb{R}^n)\colon \operatorname{spt}^\nu T\subset K\}.$$

If $\mathbb{B}$ is a solid ball and $\partial\mathbb{B}$ is its boundary, we define:

$$\mathscr{Z}_m^\nu(\mathbb{B},\partial\mathbb{B})=\{T\in\mathscr{F}_m^\nu(\mathbb{B})\colon \partial T\in\mathscr{F}_{m-1}^\nu(\partial\mathbb{B})\}$$

1.9. *Integrands.* Let U be an open subset of $\mathbb{R}^n$. A (parametric) *integrand of degree m on U* is a continuous map $\Phi\colon U\times\Lambda_m\mathbb{R}^n\to\mathbb{R}$ such that $\Phi(z,r\alpha)=r\Phi(z,\alpha)$ for all $r>0$. (Here $\Lambda_m\mathbb{R}^n$ is the set of m-vectors in $\mathbb{R}^n$.) Φ is said to be a C^k [resp. analytic] integrand in case $\Phi|U\times(\Lambda_m\mathbb{R}^n\sim\{0\})$ is C^k [resp. analytic].

Whenever $T\in\mathscr{R}_m(\mathbb{R}^n)$ and $\operatorname{spt}T\subset U$, we define:

$$\langle\Phi,T\rangle=\int\Phi[z,\vec{T}(z)]\,d\|T\|z.$$

We shall always assume that the integrand is positive and even in the sense that $\Phi[z,-\xi]=\Phi[z,\xi]$. In this case we may consider Φ to be defined on $\mathbb{R}^n\times G(n,m)$ (where $G(n,m)$ is the Grassmanian of unoriented m-plane directions in $\mathbb{R}^n$) by $\Phi(x,\Gamma)=\Phi(x,\xi)$, where ξ is any simple unit m-vector associated with Γ.

Hence for an even integrand, $\langle\Phi,T\rangle$ depends only on the measure $\|T\|$:

$$\langle\Phi,T\rangle=\int\Phi[z,\operatorname{Tan}^m(\|T\|,z)]\,d\|T\|z$$

where $\operatorname{Tan}^m(\|T\|,z)$ is the $(\|T\|,m)$ approximate tangent cone at x. (For $\|T\|$ almost all x it is the m-plane associated with $\vec{T}(x)$ [F1, 4.1.28].)

We also define for $T \in \mathscr{R}_m(\mathbb{R}^n)$ or $T \in \mathscr{R}_m^\nu(\mathbb{R}^n)$:

$$\langle \Phi, T \rangle^\nu = \int \Phi[z, \operatorname{Tan}^m(\|T\|^\nu, z)]\, d\|T\|^\nu z.$$

If $\mathbb{B}$ is a solid ball in $\mathbb{R}^n$ (or all of $\mathbb{R}^n$), we say that $R \in \mathscr{R}_m^\nu(\mathbb{B})$ is *absolutely Φ minimizing* mod ν *with respect to* $\mathbb{B}$ if and only if

$$\langle \Phi, R \rangle^\nu \leqq \langle \Phi, R + \partial T \rangle^\nu$$

whenever $T \in I_{m+1}^\nu(\mathbb{B})$. $R \in \mathscr{R}_m^\nu(\mathbb{R}^n)$ is said to be *locally Φ minimizing* mod ν if and only if $\mathbb{R}^n$ has a covering of balls $\mathbb{B}_k$ such that

$$\langle \Phi, R \rangle^\nu \leqq \langle \Phi, R + \partial T \rangle^\nu$$

whenever $T \in I_{m+1}^\nu(\mathbb{B}_k)$.

1.10. *Oriented Tangent Cones.* For $x \in \mathbb{R}^n$ and $r > 0$ *define maps* μ_r and τ_x on $\mathbb{R}^n$ by

$$\mu_r(y) = ry$$

$$\tau_x(y) = x + y.$$

If $R \in \mathscr{F}_m^{\mathrm{loc}}(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$, we define:

$$\mathrm{TAN}(R, x) = \lim_{r \to \infty} (\mu_r \circ \tau_{-x})_\# R$$

provided the limit exists (in the $\mathscr{F}$ topology on $\mathscr{F}_m^{\mathrm{loc}}(\mathbb{R}^n)$: cf. [F1, 4.3.16]). $\mathrm{TAN}(R, x)$ is, in the terminology of [F1, 4.3.16], the oriented tangent cone of R at x (provided it is unique).

Note that $\mathrm{TAN}(\partial R, x) = \partial \mathrm{TAN}(R, x)$ (provided $\mathrm{TAN}(R, x)$ exists) by the $\mathscr{F}$-continuity of ∂. Also, if $R \in \mathscr{R}_m^{\mathrm{loc}}(\mathbb{R}^n)$, then:

$$\Theta^m(\|R\|, x) = \Theta^m(\|\mathrm{TAN}(R, x)\|, 0)$$

for $\mathscr{H}^m$ almost all x. [F1, 4.3.17]

1.11. $\mathbb{E}^n$ denotes the Euclidean current on $\mathbb{R}^n$ [F1, 4.1.7, p. 357]. We define $\Omega^{n-1} \in \mathscr{R}_{n-1}(\mathbb{R}^n)$ by:

$$\Omega^{n-1} = -\partial(\mathbb{E}^n \llcorner \mathbb{B}^n).$$

Note the minus sign: Ω^{n-1} can be regarded as the unit sphere oriented *inwards*.

2. Results

2.1. Proposition. *Let $T \in \mathscr{R}_{m+1}^\nu(\mathbb{R}^{m+1})$. Then T has a unique rectifiable representative of the form $\mathbb{E}^{m+1} \llcorner t$ such that t is an integer-valued $\mathscr{L}^{m+1}$ summable function with $-\nu/2 < t(x) \leqq \nu/2$ for all $x \in \mathbb{R}^{m+1}$.*

Similarly, if $R \in \mathscr{R}_m^\nu(\partial \mathbb{B}^{m+1})$, *then* R *has a unique representative in* $\mathscr{R}_m(\partial \mathbb{B}^{m+1})$ *of the form* $\Omega^m \llcorner r$ *such that* r *is an integer-valued* $\mathscr{H}^m \llcorner \partial \mathbb{B}^{m+1}$ *summable function with support in* $\partial \mathbb{B}^{m+1}$ *with* $-\nu/2 < r(x) \leqq \nu/2$ *for all* $x \in \partial \mathbb{B}^{m+1}$.

Proof. By [F1, 4.2.26, p. 430], T has a representative mod ν of the form $\mathbb{E}^{m+1} \llcorner t$ where t is an integer-valued $\mathscr{L}^{m+1}$ summable function with $|t(x)| \leqq \nu/2$ for $x \in \mathbb{R}^{m+1}$. If ν is odd, then $|t(x)| < \nu/2$; if ν is even, then $-\nu/2 \equiv \nu/2 \pmod{\nu}$. In either case we can insist that $t(x) > -\nu/2$. Uniqueness is immediate.

The proof for $R \in \mathscr{R}_m^\nu(\partial \mathbb{B}^{m+1})$ is similar (using [F1, 4.1.15]).

2.2. *Definition.* If T [resp. R] and $\mathbb{E}^{m+1} \llcorner t$ [resp. $\Omega^m \llcorner r$] are as in Proposition (2.1) we say that $\mathbb{E}^{m+1} \llcorner t$ [resp. $\Omega^m \llcorner r$] is the *select representative* mod ν of T [resp. R].

2.3. Proposition. *Let* $T_1, T_2 \in \mathscr{R}_{m+1}(\mathbb{R}^{m+1})$ *be select representatives* mod ν. *Then:*

$$\mathbf{M}(T_1 - T_2) \leqq (\nu - 1)\, \mathbf{M}^\nu(T_1 - T_2) \tag{1}$$

$$\mathbf{M}(\partial T_1) \qquad \leqq (\nu - 1)\, \mathbf{M}^\nu(\partial T_1). \tag{2}$$

Proof. Writing $T_i = \mathbb{E}^{m+1} \llcorner t_i$ as above, we see that $|t_1(x) - t_2(x)| \leqq \nu - 1$ for all x. Hence:

$$|t_1(x) - t_2(x)| \leqq (\nu - 1)\, |(t_1(x) - t_2(x)) \pmod{\nu}|.$$

This proves (1).

We prove (2) first when T_1 is a polyhedral chain. Let R be a polyhedron in ∂T_1. If t_1 takes the value i on one side of R and j on the other, then R has multiplicity $|i - j| \leqq \nu - 1$ in ∂T_1. Thus:

$$|i - j| \leqq (\nu - 1)\, |(i - j) \pmod{\nu}|.$$

This proves (2) for polyhedral chains.

Now let $T \in \mathscr{R}_{m+1}(\mathbb{R}^{m+1})$ be any select representative mod ν. By [F1, 4.2.22$^\nu$] there exists a sequence of polyhedral chains P_j such that:

$$\lim \mathscr{F}^\nu(P_j - T) = 0 \tag{3}$$

$$\lim \mathbf{M}^\nu(\partial P_j) = \mathbf{M}^\nu(\partial T). \tag{4}$$

Since $\mathscr{F}^\nu = \mathbf{M}^\nu$ on $m+1$ chains in $\mathbb{R}^{m+1}$, (3) implies:

$$\lim \mathbf{M}^\nu(P_j - T) = 0. \tag{5}$$

Note that we can choose the P_j to be select representatives mod ν. Then by (1) and (5):

$$\lim \mathbf{M}(P_j - T) = 0.$$

Hence:

$$\lim \mathscr{F}(\partial P_j - \partial T) = 0.$$

By lower semicontinuity of $\mathbf{M}$ with respect to $\mathscr{F}$ and by (2) for the polyhedral chains P_j:

$$\begin{aligned}\mathbf{M}(\partial T) &\leqq \liminf \mathbf{M}(\partial P_j)\\ &\leqq \liminf (v-1)\,\mathbf{M}^v(\partial P_j)\\ &= (v-1)\,\mathbf{M}^v(\partial T)\end{aligned}$$

by (4).

Corollary. *Let* $T \in \mathscr{R}^v_{m+1}(\mathbb{R}^{m+1})$. *Then* $T \in I^v_{m+1}(\mathbb{R}^{m+1})$ *if and only if its select representative* mod v *is an integral current.*

Proof. The result follows immediately from 2.3(2) and [F1, 4.2.16 and 4.2.16^v].

2.4. Proposition. *Let* $T = (\mathbb{E}^{m+1} \llcorner t) \in I_{m+1}(\mathbb{B}^{m+1})$ *and* $R = (\Omega^m \llcorner r) \in \mathscr{R}_m(\partial \mathbb{B}^{m+1})$ *be select representatives* mod 4. *For* $i = -1, 1$ *define*

$$T_i = \mathbb{E}^{m+1} \llcorner \chi_{i,2}(t)$$

$$R_i = \Omega^m \llcorner \chi_{i,2}(r)$$

where $\chi_{i,2}$ *is the characteristic function of* $\{i, 2\}$.

Then T_i, R_i *are select representatives* mod 2 *and:*

1) $\|R + \partial T\|^{(4)} = \sum_{i=-1,1} \|R_i + \partial T_i\|^{(2)}$.

Proof. By definition, T_i and R_i are select representatives mod 2. By the Gauss-Green Theorem [F1, 4.5.6], there is a set N with $\mathscr{H}^m(N) = 0$ such that for each $x \in \mathbb{R}^{m+1} \sim N$, there is an oriented m-dimensional hyperplane Γ_x such that:

2) $\mathrm{TAN}(T, x) = \mathbb{E}^{m+1} \llcorner \tau$

where $\tau = j$ on one side of Γ_x and $\tau = k$ on the other side, and:

3) $\mathrm{TAN}(\partial T, x) = (j-k)\Gamma_x$.

If also $x \in \partial \mathbb{B}^{m+1}$, then:

4) $\Gamma_x = -\mathrm{TAN}(\Omega^m, x)$ and $k = 0$.

Note that 2) implies

$$\mathrm{TAN}(T_i, x) = \mathbb{E}^{m+1} \llcorner \chi_{i,2}(\tau).$$

Hence:

5) $\mathrm{TAN}(\partial T_i, x) = (\chi_{i,2}(j) - \chi_{i,2}(k))\Gamma_x$.

Also, since $R = \Omega^m \llcorner r$ is rectifiable, there is a set N' with $\mathscr{H}^m(N') = 0$ such that for $x \in \partial \mathbb{B}^{m+1} \sim N'$:

6) $\mathrm{TAN}(R, x) = r(x)\,\mathrm{TAN}(\Omega^m, x)$.

Hence:

7) $\mathrm{TAN}(R_i, x) = \chi_{i,2}(r(x))\,\mathrm{TAN}(\Omega^m, x)$.

By 1.7(2) it suffices to show that for $\mathscr{H}^m$ almost all x:

$$\Theta^m(\|R+\partial T\|^{(4)}, x) = \sum_{i=-1,1} \Theta^m(\|R_i+\partial T_i\|^{(2)}, x)$$

or (1.7(1)):

$$|\Theta^m(\|R+\partial T\|, x)(\mathrm{mod}\,4)| = \sum_{i=-1,1} |\Theta^m(\|R_i+\partial T_i\|, x)(\mathrm{mod}\,2)|.$$

For $\mathscr{H}^m$ almost all x, this is equivalent (1.10) to:

8) $|\Theta^m(\|\mathrm{TAN}(R+\partial T, x)\|, 0)(\mathrm{mod}\,4)|$
$$= \sum_{i=-1,1} |\Theta^m(\|\mathrm{TAN}(R_i+\partial T_i, x)\|, 0)(\mathrm{mod}\,2)|.$$

We shall show 8) for all $x \in \mathbb{R}^{m+1} \sim (N \cup N')$.

Case 1. $x \notin \partial \mathbb{B}^{m+1}$. Then $x \notin \operatorname{spt} R$, so:

9) $|\Theta^m(\|\mathrm{TAN}(R+\partial T, x)\|, 0)(\mathrm{mod}\,4)|$
$$= |\Theta^m(\|\mathrm{TAN}(\partial T, x)\|, 0)(\mathrm{mod}\,4)|$$
$$= |(j-k)(\mathrm{mod}\,4)|$$
by 3).

Also:

10) $|\Theta^m(\|\mathrm{TAN}(R_i+\partial T_i, x)\|, 0)(\mathrm{mod}\,2)|$
$$= |\Theta^m(\|\mathrm{TAN}(\partial T_i, x)\|, 0)(\mathrm{mod}\,2)|$$
$$= |(\chi_{i,2}(j) - \chi_{i,2}(k))(\mathrm{mod}\,2)|$$
by 5).

Now 8) follows from 9), 10), and the combinatorial equation:

11) $|(j-k)(\mathrm{mod}\,4)| = \sum_{i=-1,1} |(\chi_{i,2}(j) - \chi_{i,2}(k))(\mathrm{mod}\,2)|$

which holds for $j, k \in \{-1, 0, 1, 2\}$.

Case 2. $x \in \partial \mathbb{B}^{m+1}$. Then by 3), 4), and 6):

$$\mathrm{TAN}(R+\partial T, x) = \mathrm{TAN}(R, x) + \mathrm{TAN}(\partial T, x)$$
$$= (j - r(x))\Gamma_x$$

12) $\therefore |\Theta^m(\|\mathrm{TAN}(R+\partial T, x)\|, 0)(\mathrm{mod}\,4)| = |(j-r(x))(\mathrm{mod}\,4)|$.

Likewise by 4), 5), and 7):

$$\mathrm{TAN}(R_i+\partial T_i x) = \mathrm{TAN}(R_i, x) + \mathrm{TAN}(\partial T_i, x)$$
$$= (\chi_{i,2}(j) - \chi_{i,2}(r(x)))\Gamma_x$$

so

13) $|\Theta^m(\|\mathrm{TAN}(R_i+\partial T_i, x)\|, 0)(\mathrm{mod}\, 2)| = |(\chi_{i,2}(j)-\chi_{i,2}(r(x)))(\mathrm{mod}\, 2)|$.

Note that 8) follows from 12), 13), and 11) (setting $k=r(x)$).

2.5. *Definition.* If $T\in I^4_{m+1}(\mathbb{B}^{m+1})$ and $\mathbb{E}^{m+1} \llcorner t$ is its select representative, define for $i=1, -1$:

$$\psi_i(T) = (\mathbb{E}^{m+1} \llcorner \chi_{i,2}(t))(\mathrm{mod}\, 2).$$

Similarly, if $R\in \mathscr{R}^4_m(\partial \mathbb{B}^{m+1})$ and $\Omega^m \llcorner r$ is its select representative, define for $i=1, -1$:

$$\varphi_i(R) = (\Omega^m \llcorner \chi_{i,2}(r))(\mathrm{mod}\, 2)$$

2.6. Proposition. *The maps*

$$\psi_1 \times \psi_{-1}\colon\ I^4_{m+1}(\mathbb{B}^{m+1}) \to I^2_{m+1}(\mathbb{B}^{m+1}) \times I^2_{m+1}(\mathbb{B}^{m+1})$$

and

$$\varphi_1 \times \varphi_{-1}\colon\ \mathscr{R}^4_m(\partial\mathbb{B}^{m+1}) \to \mathscr{R}^2_m(\partial\mathbb{B}^{m+1}) \times \mathscr{R}^2_m(\partial\mathbb{B}^{m+1})$$

are bijections and

$$\|R+\partial T\|^4 = \sum_{i=1,-1} \|\varphi_i(R)+\partial\psi_i(T)\|^2$$

for all $R\in\mathscr{R}^4_m(\partial\mathbb{B}^{m+1})$ *and* $T\in I^4_{m+1}(\mathbb{B}^{m+1})$.

Proof. If $T\in I^4_{m+1}(\mathbb{B}^{m+1})$ and $\mathbb{E}^{m+1}\llcorner t$ is its select representative, then by corollary (2.3), $\mathbb{E}^{m+1}\llcorner t$ is an integral current, so $\psi_i(T)\in I^2_{m+1}(\mathbb{B}^{m+1})$. That is, ψ_i does in fact map $I^4_{m+1}(\mathbb{R}^{m+1})$ to $I^2_{m+1}(\mathbb{B}^{m+1})$.

Note that $t\mapsto\langle\chi_{1,2}(t),\ \chi_{-1,2}(t)\rangle$ is a bijection from $\mathbb{Z}^4$ to $\mathbb{Z}^2\times\mathbb{Z}^2$. This induces a bijection

$$\mathbb{E}^{m+1}\llcorner t \to \langle \mathbb{E}^{m+1}\llcorner\chi_{1,2}(t),\ \mathbb{E}^{m+1}\llcorner\chi_{-1,2}(t)\rangle$$

of select representatives mod 4 and pairs of select representatives mod 2. This in turn induces a bijection of $I^4_{m+1}(\mathbb{B}^{m+1})$ and $I^2_{m+1}(\mathbb{B}^{m+1})\times I^2_{m+1}(\mathbb{B}^{m+1})$. This bijection is clearly $\psi_1\times\psi_{-1}$. Similarly $\varphi_1\times\varphi_{-1}$ is a bijection.

The rest of the proposition follows immediately from Proposition (2.4).

2.7. Theorem. *Let Φ be an even integrand of degree m on $\mathbb{B}^{m+1}$. Let $R\in\mathscr{R}^4_m(\partial\mathbb{B}^{m+1})$ and $T\in I^4_{m+1}(\mathbb{B}^{m+1})$. Then $R+\partial T$ is absolutely Φ minimizing* mod 4 with respect to $\mathbb{B}^{m+1}$ *if and only if $\varphi_i(R)+\partial\psi_i(T)$ is absolutely Φ minimizing* mod 2 *with respect to* $\mathbb{B}^{m+1}$ *for* $i=-1, 1$.

Proof. Suppose $R+\partial T$ is not absolutely Φ minimizing mod 4 with respect to $\mathbb{B}^{m+1}$. Then there is a $T'\in I^4_{m+1}(\mathbb{B}^{m+1})$ such that:

$$\langle\Phi,\ R+\partial T'\rangle^4 < \langle\Phi, R+\partial T\rangle^4.$$

But then by Proposition 2.6:

$$\sum_{i=-1,1} \langle \Phi, \varphi_i(R)+\partial\psi_i(T')\rangle^2 < \sum_{i=-1,1} \langle \Phi, \varphi_i(R)+\partial\psi_i(T)\rangle^2$$

so at least one of the $\varphi_i(R)+\partial\psi_i(T)$ is not absolutely Φ minimizing mod 2 with respect to $\mathbb{B}^{m+1}$.

Conversely if, say, $\varphi_{-1}(R)+\partial\psi_{-1}(T)$ is not absolutely Φ minimizing mod 2 with repsect to $\mathbb{B}^{m+1}$, then there is a $T_{-1}\in I^2_{m+1}(\mathbb{B}^{m+1})$ such that:

$$\langle \Phi, \varphi_{-1}(R)+\partial T_{-1}\rangle^2 < \langle \Phi, \varphi_{-1}(R)+\partial\psi_{-1}(T)\rangle^2$$

$$\therefore \langle \Phi, \varphi_1(R)+\partial\psi_1(T)\rangle^2 + \langle \Phi, \varphi_{-1}(R)+\partial T_{-1}\rangle^2 < \sum_{i=-1,1} \langle \Phi, \varphi_i(R)+\partial\psi_i(T)\rangle^2.$$

If $T'=(\psi_1\times\psi_{-1})\langle\psi_1(T), T_{-1}\rangle$, the last inequality becomes by Proposition 2.6:

$$\langle \Phi, R+\partial T'\rangle^4 < \langle \Phi, R+\partial T\rangle^4$$

whence $R+\partial T$ is not absolutely Φ minimizing mod 4 with respect to $\mathbb{B}^{m+1}$.

Corollary. *If* $Q\in\mathscr{Z}^4_m(\mathbb{B}^{m+1}, \partial\mathbb{B}^{m+1})\cap\mathscr{R}^4_m(\mathbb{B}^{m+1})$ *is absolutely* Φ *minimizing* mod 4 *with respect to* $\mathbb{B}^{m+1}$, *then* $\operatorname{spt} Q=\bigcup \operatorname{spt} Q_i$ $(i=-1,1)$, *where* $Q_i\in\mathscr{Z}^2_m(\mathbb{B}^{m+1}, \partial\mathbb{B}^{m+1})\cap\mathscr{R}^2_m(\mathbb{B}^{m+1})$ *is absolutely* Φ *minimizing* mod 2 *with respect to* $\mathbb{B}^{m+1}$ *for* $i=1, -1$.

Conversely, if $Q_i\in\mathscr{Z}^2_m(\mathbb{B}^{m+1}, \partial\mathbb{B}^{m+1})\cap\mathscr{R}^2_m(\mathbb{B}^{m+1})$ *is absolutely* Φ *minimizing* mod 2 *with respect to* $\mathbb{B}^{m+1}$ (for $i=-1,1$), *then there is a* $Q\in\mathscr{Z}^4_m(\mathbb{B}^{m+1}, \partial\mathbb{B}^{m+1})\cap\mathscr{R}^4_m(\mathbb{B}^{m+1})$, *which is absolutely* Φ *minimizing* mod 4 *with respect to* $\mathbb{B}^{m+1}$, *such that* $\operatorname{spt} Q=\bigcup \operatorname{spt} Q_i$ $(i=1, -1)$. (The corollary follows because every $Q\in\mathscr{Z}^\nu_m(\mathbb{B}^{m+1}, \partial\mathbb{B}^{m+1})\cap\mathscr{R}^\nu_m(\mathbb{B}^{m+1})$ is of the form $R+\partial T$ for some $R\in\mathscr{R}^\nu_m(\partial\mathbb{B}^{m+1})$ and $T\in I^\nu_{m+1}(\mathbb{B}^{m+1})$.)

2.8. Lemma. *Let* Φ *be an analytic parametric elliptic integrand of degree* m *on* $\mathbb{R}^{m+1}$ *(or more generally on an open subset of* $\mathbb{R}^{m+1}$*). Let* F: $\mathbb{R}^m\times\mathbb{R}\times\mathbb{R}^m\to\mathbb{R}$ *be the associated nonparametric integrand, where for all* C^1 *functions* f: $\mathbb{B}^m(0,1)\to\mathbb{R}$ *one sets:*

$$F(f)=\int_{x\in\mathbb{B}^m(0,1)} F(x,f(x),Df(x))\,d\mathscr{L}^m(x).$$

Let f,g: $\mathbb{B}^m(0,1)\to\mathbb{R}$ *be distinct analytic functions which are stationary for* F. *Suppose* $f(0)=g(0)$ *and* $Df(0)=Dg(0)=0$ *and let:*

$$f(x)-g(x)=\sum_{k=n}^{\infty} C_k(x)$$

be the Taylor series for $f-g$ *about* 0, *where* C_k is a homogeneous polynomial of degree k and $C_n\neq 0$.

Then there is a linear isomorphism M: $\mathbb{R}^m\to\mathbb{R}^m$ *such that* $C_n(Mx)$ *is harmonic. If* Φ *is the area integrand or an analytic function of* $\mathbb{R}^{m+1}$ *times the area integrand, then* $C_n(x)$ *is harmonic.*

Proof. The Euler-Lagrange equations for f are (in the usual notation) of the form:

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial^2 F}{\partial p_i\,\partial p_j}(x,f(x),Df(x))\cdot\frac{\partial^2 f}{\partial x_i\,\partial x_j}(x)=u(x,f(x),Df(x))$$

where u is analytic. Subtracting the corresponding equation for g:

$$\begin{aligned}&\sum_{i,j}\frac{\partial^2 F}{\partial p_i\,\partial p_j}(x,f(x),Df(x))\frac{\partial^2(f-g)}{\partial x_i\,\partial x_j}(x)\\&+\sum_{i,j}\left\{\frac{\partial^2 F}{\partial p_i\,\partial p_j}(x,f(x),Df(x))-\frac{\partial^2 F}{\partial p_i\,\partial p_j}(x,g(x),Dg(x))\right\}\frac{\partial^2 g}{\partial x_i\,\partial x_j}(x)\\&=u(x,f(x),Df(x))-u(x,g(x),Dg(x)).\end{aligned}$$

Since $f-g$ has a zero of order n at 0, the right-hand side and the second term of the left-hand side above have a zero of order at least $n-1$. Thus:

$$\sum_{i,j}\frac{\partial^2 F}{\partial p_i\,\partial p_j}(x,f(x),Df(x))\cdot\frac{\partial^2(f-g)}{\partial x_i\,\partial x_j}(x)=\mathcal{O}(|x|^{n-1}).$$

Let A_{ij} be the constant term in the Taylor series of $(\partial^2 F/\partial p_i\,\partial p_j)(x,f(x),Df(x))$ about 0. Then:

$$\sum_{i,j}A_{ij}\frac{\partial^2(f-g)}{\partial x_i\,\partial x_j}(x)+\sum_{i,j}h(x)\frac{\partial^2(f-g)}{\partial x_i\,\partial x_j}(x)=\mathcal{O}(|x|^{n-1})$$

where $A=[A_{ij}]$ is a positive definite matrix (by ellipticity of Φ) and $h(0)=0$. Hence the second term on the left is also $\mathcal{O}(|x|^{n-1})$:

$$\sum_{i,j}A_{ij}\frac{\partial^2(f-g)}{\partial x_i\,\partial x_j}(x)=\mathcal{O}(|x|^{n-1})$$

$$\therefore\sum_{i,j}A_{ij}\frac{\partial^2}{\partial x_i\,\partial x_j}C_n(x)=\mathcal{O}(|x|^{n-1}).$$

But C_n is a homogeneous polynomial of degree n; if the left-hand side did not vanish, it would be of degree $n-2$. Therefore:

$$\sum_{i,j}A_{ij}\frac{\partial^2}{\partial x_i\,\partial x_j}C_n(x)=0.$$

Since $A=[A_{ij}]$ is positive definite, there is a symmetric nonsingular matrix M such that $M^2=A$. Thus

$$\sum_{i=1}^{m}\left(\frac{\partial}{\partial x_i}\right)^2 C_n(Mx)=\sum_{i=1}^{m}\sum_{j=1}^{m}A_{ij}\frac{\partial^2}{\partial x_i\,\partial x_j}C_n(x)=0.$$

Consequently $C_n(Mx)$ is harmonic. In case Φ is the area integrand or an analytic function of $\mathbb{R}^{m+1}$ times the area integrand, then A is the identity matrix, so $C_n(x)$ is harmonic.

Corollary. *Let f, g, and M be as above. Set $D=\{x\in U^m(0,1)=f(x)=g(x)\}$. Then:*

$$M^{-1}\operatorname{Tan}(D,0)=\operatorname{Tan}(M^{-1}D,0)$$

is the zero set of a homogeneous harmonic polynomial.

Remark. In case $m=2$, one can show by writing $C_n(Mx)$ in polar coordinates that, in a sufficiently small neighborhood of 0, the set D actually consists of n analytic curves passing through 0 [N, § 437, p. 399]. When $m>2$, I do not know whether D and $\operatorname{Tan}(D,0)$ need be homeomorphic near 0.

2.9. Theorem. *Let Φ be a positive even parametric elliptic integrand of degree m and class $q+1\geqq 3$ on $\mathbb{R}^{m+1}$. Let $Q\in\mathscr{R}_m^4(\mathbb{R}^{m+1})$ be locally Φ minimizing* mod 4. *Then:*

1) *There exists a closed set N with $\mathscr{H}^{m-2}(N\sim\operatorname{spt}\partial Q)=0$ such that $\operatorname{spt}Q\sim N$ is an immersed C^q manifold with no points of multiplicity >2. Locally, each sheet of $\operatorname{spt}Q\sim N$ is Φ minimizing. Let D be the set of double points of $\operatorname{spt}Q\sim N$, and let Σ be the set of double points at which the two sheets are tangent.*

2) *If Φ is analytic, then $\operatorname{spt}Q\sim N$ is an immersed analytic manifold, $D\sim\Sigma$ is an analytic $(m-1)$ dimensional manifold, and Σ is an analytic variety of dimension $\leqq m-2$. For all $x\in\Sigma$, $\operatorname{Tan}(D,x)$ is the image under a linear monomorphism $\mathbb{R}^m\to\mathbb{R}^{m+1}$ of the zero set of a homogeneous harmonic polynomial of degree $\geqq 2$.*

3) *If Φ is an analytic function of $\mathbb{R}^{m+1}$ times the area integrand, then the linear monomorphism of* 2) *is an isometry.*

4) *If Φ is the area integrand, then the Hausdorff dimension of $N\sim\operatorname{spt}Q$ is $\leqq m-7$.*

Corollary. *Let $Q\in\mathscr{R}_2^4(\mathbb{R}^3)$ be locally area minimizing. Then $\operatorname{spt}Q\sim\operatorname{spt}\partial Q$ is an immersed analytic manifold. Its self-intersection set consists of analytic curves along which two sheets cross transversely, together with isolated points where two sheets are tangent and at which an even number of such curves meet at equal angles.*

Proof 1) *and* 4). Let $\{\mathbb{B}_k\}$ be a locally finite set of closed balls such that $\bigcup(\operatorname{int}\mathbb{B}_k)=\mathbb{R}^{m+1}\sim\operatorname{spt}\partial Q$ and such that each $Q\llcorner\mathbb{B}_k$ is absolutely Φ minimizing mod 4 with repsect to $\mathbb{B}_k$. Then by the corollary to Theorem 2.7, $\operatorname{spt}(Q\llcorner\mathbb{B}_k)=\bigcup_{i=-1,1}\operatorname{spt}Q_{i,k}$ where $Q_{i,k}\in\mathscr{Z}_m^2(\mathbb{B}_k,\partial\mathbb{B}_k)\cap\mathscr{R}_m^2(\mathbb{B}_k)$ is absolutely Φ minimizing mod 2 with respect to $\mathbb{B}_k$.

By [ASS], the singular set $N_{i,k}$ of $Q_{i,k}\llcorner\operatorname{int}\mathbb{B}_k$ is closed in $\operatorname{int}\mathbb{B}_k$ and $\mathscr{H}^{m-2}(N_{i,k})=0$. Furthermore, if Φ is the area integrand, then the Hausdorff dimension of $N_{i,k}$ is $\leqq m-7$ [F2]. Letting $N=(\bigcup N_{i,k})\cup\operatorname{spt}\partial Q$ gives 1) and 4).

2) and 3). Suppose Φ is analytic. Then $D\sim\Sigma$ is (locally) the intersection of two m-dimensional analytic manifolds which cross transversely and hence is an $(m-1)$ dimensional analytic manifold.

Clearly D is an analytic variety, so by [G] it can be triangulated analytically. Hence $\dim D\leqq m-1$, since otherwise the two sheets of $\operatorname{spt}Q\sim N$ would, by

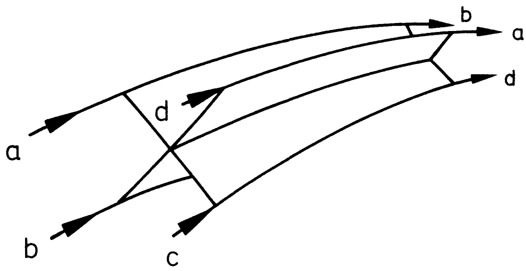


Fig. 4. An area minimizing flat chain mod 4 which decomposes locally, but not globally, into two minimal surfaces

analyticity, coincide (constituting a single imbedded manifold with multiplicity 2 in Q). Now by Lemma 2.8, there is *no* point $x \in \Sigma$ such that $\mathrm{Tan}(D, x)$ is an $(m-1)$ subspace of $\mathbb{R}^{m+1}$. Hence each $x \in \Sigma$ must lie in the $(m-2)$ skeleton of any triangulation of D. Consequently Σ must have dimension $\leqq m-2$.

The rest of 2), as well as 3), is just lemma 2.8.

2.10. Global Behavior. We have seen that any Φ minimizing flat chain mod 4 decomposes locally into two Φ minimizing surfaces. Globally, however, it need not. Consider, for instance, a simple closed curve which winds four times around a torus. If the torus is sufficiently thin, the area minimizing flat chain mod 4 will lie within the torus as in Fig. 4. Locally it looks like two intersecting ribbons, but globally the "two" are one (an immersion of a Möbius strip in $\mathbb{R}^3$).

We remark that Theorem 2.9 extends in an obvious way to arbitrary manifolds.

References

[ASS] Almgren, F.J. Jr., Schoen, R., Simon, L.: Regularity and singularity estimates on hypersurfaces minimizing parametric elliptic variational integrals. Acta Mathematica, **139**, 217–265 (1977)

[F1] Federer, H.: Geometric Measure Theory. Berlin-Heidelberg-New York: Springer 1969

[F2] Federer, H.: The singular sets of area minimizing rectifiable currents with codimension one and of area minimizing flat chains modulo two with arbitrary codimension. Bull. Am. Math. Soc., **76**, 767–771 (1970)

[G] Giesecke, B.: Simpliziale Zerlegungsabzähler analytischer Räume. Math. Zeitschr., **83**, 177–213 (1964)

[N] Nitsche, J.C.C.: Vorlesungen über Minimalflächen. Berlin-Heidelberg-New York: Springer 1975

[T] Taylor, J.E.: Regularity of the singular sets of two-dimensional area-minimizing flat chains modulo 3 in $\mathbb{R}^3$. Invent. Math., **22**, 119–159 (1973)

Received November 2, 1978

Inventiones math. 53, 59–75 (1979)

Inventiones mathematicae

Absolute Stability and Bifurcation Theory

Ricardo Mañé
I.M.P.A., Rio de Janeiro, Instituto de Matemática U.F.R.G.S., Porto Alegre, Brasil

Bifurcation theory studies families of diffeomorphisms of a compact smooth boundaryless manifold M modelled on some compact smooth manifold V, i.e. C^k maps $f\colon V \to \mathrm{Diff}^r(M)$ where $\mathrm{Diff}^r(M)$ denotes the Banach manifold of C^r diffeomorphisms of M.

One of the natural objectives of this theory is to find elements of the family exhibiting interesting persistent properties, more precisely values $t \in V$ such that certain features of the orbit structure of f_t can also be found in the orbit structure of $g_{\bar{t}}$, with $\bar{t}$ near to t, for every nearby family g. There exist several ways of formalizing this idea (see Newhouse, Palis, Takens [10] or Robinson [14] for instance). We shall consider the following one: let $F^{k,r}(V, M)$ be the Banach manifold of C^k maps of V in $\mathrm{Diff}^r(M)$. We say that $f \in F^{k,r}(V, M)$ is *stable* at $t \in V$ if for all $\varepsilon > 0$ there exists a neighborhood $\mathscr{U}$ of f in $F^{k,r}(V, M)$ such that if $g \in \mathscr{U}$ there exists $\bar{t} \in V$ with $d(\bar{t}, t) \leqq \varepsilon$ and a homeomorphism $h\colon M \hookleftarrow$ such that $gh = hf$.

It follows from the results of Sotomayor [18], Brunovsky [2], and well known properties of rotation numbers that generic elements of $F^{k,r}([0,1], S^1)$ (with $k \geqq 1\, r \geqq 2$) are stable at every value of the parameter. This property doesn't extend to higher dimensions. Using the methods of Robinson and Williams [16], or Palis [13], it is possible to find, given V, $k \geqq 1$, $r \geqq 1$ and M with $\dim M \geqq 2$, an open set $\mathscr{U} \subset F^{k,r}(V, M)$ such that f is unstable at every value of the parameter for all $f \in \mathscr{U}$.

The characterization of the stable values of the parameter is a beautiful objective but, in that generality, out of seope for the present state of development of the theory since easier problems, like the characterization of structurally, stable diffeomorphisms, are far from being solved. About this last problem much has been done starting from an idea of Franks [5], later developed in [6] and [11], consisting in replacing structural stability by a stronger and easier to handle property. In our context the corresponding definition would be the following: $f \in F^{k,r}(V, M)$ is *absolutely stable* at $t \in V$ if

0020-9910/79/0053/0059/$03.40

there exist a neighborhood $\mathscr{U}$ of f in $F^{k,r}(V, M)$ and a constant $k_1 > 0$ such that for all $g \in \mathscr{U}$ there exist $\bar{t} \in V$ and a homeomorphism h: $M \hookleftarrow$ such that:

a) $hg = fh$,

b) $d(\bar{t}, t) \leqq k_1 d_0(g, f)$, where $d_0(g,f) = \sup_{s \in V} (\sup_{x \in M} d(f_s(x), g_s(x)))$,

c) $d_0(h, I) \leqq k_1 d_0(g, f)$ where $d_0(h, I) = \sup_{x \in M} d(h(x), x)$.

The next theorem will show that this concept is too restrictive to be useful since no real bifurcation may occur at absolutely stable values of the parameter.

Theorem A. *If $f \in F^{k,r}(V, M)$ is absolutely stable at $t \in V$ then f_t satisfies Axiom A and the strong transversality condition.*

However the situation is completely different if instead of requiring the conjugacy h to be globally defined we ask only for a conjugacy between certain interesting invariant sets, for instance between the non wandering sets $\Omega(g_{\bar{t}})$, $\Omega(f_t)$ [5]. For reasons that will be discussed later instead of using the non wandering set we shall prefer the set of *chain recurrent* points.

Definition [4]. *If φ is a homeomorphism of the compact metric space K we shall say that $x \in K$ is chain recurrent if for all $\varepsilon > 0$ there exist points $x = x_0, x_1, \ldots, x_n = x$ satisfying $d(f(x_j), x_{j+1}) \leqq \varepsilon$ for all $a \leqq j \leqq n$. The set of chain recurrent points will be denoted $\Gamma(\varphi)$.*

For properties about this set see [1, 4] and [17]. It is interesting to remark that for Axiom A diffeomorphisms the set of chain recurrent point coincides with the set of points that belong to a cycle (for the definition of cycle see [5]). Moreover in [4] it was proved that a diffeomorphism satisfies Axiom A and the no cycles property if and only if its Γ set is hyperbolic.

Another concept that will be useful is that of AS sets of a diffeomorphism.

Definition. *A compact invariant set Λ of a diffeomorphism $f \in \mathrm{Diff}^1(M)$ is an AS set if $\Gamma(f/\Lambda)$ is a hyperbolic set with local product structure* [7], *the periodic points of f/Λ are dense in $\Gamma(f/\Lambda)$ and for every $x \in \Lambda$, if $x \in W^s(x_1) \cap W^u(x_2)$ with $x_1 \in \Gamma(f/\Lambda)$, $x_2 \in \Gamma(f/\Lambda)$ then $W^s(x_1)$ and $W^u(x_2)$ intersect transversally at x. Moreover we say that Λ is an isolated AS set if for some compact neighborhood U of Λ we have $\Lambda = \bigcap_n f^n(U)$.*

In other words Λ is an AS set of a diffeomorphism f if f satisfies Axiom A and the strong transversality condition "at Λ". Observe that a diffeomorphism satisfies Axiom A and the strong transversality condition if and only if the whole manifold is an AS set.

In this definition we denoted by $W^s(x)$, $W^u(x)$ the stable and unstable manifolds of $x \in \Gamma(f/\Lambda)$ defined as usual as the sets of points asymptotic with x in the past or future respectively. In [15] Robinson proved that if Λ is an isolated AS set of f and U is a compact neighborhood of Λ satisfying $\bigcap_n f^n(U) = \Lambda$ then for every g near to f in the C^1 topology there exists a homeomorphism h: $\Lambda \to \bigcap_n g^n(U)$ such that $gh(x) = hf(x)$ for all $x \in \Lambda$.

The last definition required for the statement of the next theorem is that of *non transversal* orbits.

Definition. *An orbit* γ *of* $f \in \mathrm{Diff}^r(M)$ *is non transversal if* $\omega(\gamma)$ *and* $\alpha(\gamma)$ *are contained in hyperbolic sets* Λ_1, Λ_2 *and there exist* $y_1 \in \Lambda_1$, $y_2 \in \Lambda_2$ *satisfying*

$$\lim_{n \to +\infty} d(f^n(x), f^n(y_1)) = \lim_{n \to +\infty} d(f^{-n}(x), f^{-n}(y_2)) = 0,$$

such that $W^s(y_1)$ *and* $W^s(y_2)$ *meet non transversally at* x.

Now we are ready to state the following result giving necessary conditions for the absolute Γ-stability of a family of diffeomorphisms at a certain value of the parameter. Absolute Γ-stability is defined as follows: $f \in F^{k,r}(V, M)$ is absolutely Γ-stable at $t \in V$ if there exist a constant $k>0$ and a neighborhood $\mathcal{U}$ of f in $F^{k,r}(V, M)$ such that for all $g \in \mathcal{U}$ there exist $\bar{t} \in V$ and a homeomorphism $h\colon \Gamma(f_t) \to \Gamma(g_{\bar{t}})$ satisfying:

1) $gh(x) = hf(x)$ for all $x \in \Gamma(f_t)$,
2) $d(\bar{t}, t) \leqq k d_0(g, f)$,
3) $d_0(h, I) \leqq k d_0(g, f)$,

where $d_0(g,f)$ is defined as before and $d_0(h, I) = \sup\{d(h(x), x) \mid x \in \Gamma(f_t)\}$. In this case we shall also say that $t \in V$ is an absolutely Γ-stable value of the parameter.

Theorem B. *If* $f \in F^{k,r}(V, M)$ *is absolutely* Γ*-stable at* $t \in V$ *then* $\Gamma(f_t)$ *is a finite disjoint union of a finite set* $\beta_1, \ldots, \beta_m$ *of non hyperbolic periodic orbits, an isolated AS set* Γ_0 *and a finite set of non transversal orbits* $\gamma_1, \ldots, \gamma_k$ *with* $\alpha(\gamma_i) \cup \omega(y_i) \subset \Gamma_0$ *for every* $1 \leqq i \leqq k$. *Moreover if* $\Gamma(f_t)$ *doesn't contain non transversal orbits then* $\Gamma(f_t/\Gamma_0) = \Gamma_0$.

This conditions are clearly unsufficient to grant the absolute Γ-stability of f at V. Since there are no Γ-explosions (i.e. for any diffeomorphism f given a neighborhood U of $\Gamma(f)$ then $\Gamma(g) \subset U$ for every g C^0 near to f) and by the stability of the AS set, what is missing in order to have a characterization of the absolutely Γ-stable values of the parameter, are conditions ensuing the absolute stability of the non transversal and non hyperbolic orbits. It looks a reasonable and interesting problem to complete this characterization at least for low dimensional V's. When $\dim V \geqq 3$, difficult problems may arise because of the non hyperbolic periodic points (see [19] for an analysis of the singularities that can appear in a generic family of flows). It would be interesting to characterize the absolutely Γ-stable values of the parameter by a transversality property in the following sense: given $m \in \mathbb{Z}^+$ find immerse submanifolds $V_0, V_1, \ldots, V_l$ such that $V_i \cap (\bigcup_{j<i} V_j) = \emptyset$ and V_0 is the set of diffeomorphisms satisfying Axiom A and the no cycles condition, such that $f \in F^{k,r}(V, M)$ is absolutely Γ-stable at $t \in V$ if $f_t \in \bigcup_j V_j$ and $f(V)$ intersects $\bigcup_j V_j$ transversally at f_t.

We had two main reasons for choosing the Γ set instead of the more traditional non wandering set for our definition. The first is that we weren't able to prove an analogous result for absolutely Ω-stable vaues of the parameter. We could prove only that $\Omega(f_t)$ would be contained in (but not that coincides with) a set with a decomposition like that given in Theorem B for $\Gamma(f_t)$. Second, even having a more complete result for the structure of $\Omega(f_t)$ the possibilities of developing necessary and sufficient conditions for absolute Ω-stability seem small without an analysis of $\Gamma(f_t)$ because this is the set of "cycles" that could be reached by Ω-explosions. Since the analysis of cycles are one of the main goals of bifurcation theory (see Newhouse, Palis [9] for general comments on this subject) it seems natural to work with a set that includes this phenomena.

Examples of absolute Γ-stability can be found in [8] or [9]. Problem 1 in [8] p. 312 amounts to ask whether generic one parameter families of diffeomorphisms starting at a Morse-Smale are absolutely Γ-stable at its first bifurcation point. Finally, Dankner's example [3] satisfies the necessary conditions of Theorem B so its seems natural to ask if there exists $f \in F^{k,r}(D^n, M^3)$ (where D^n denotes the n-disc) absolutely Γ-stable at $t=0$ and with f_0 being Dankner's example.

The methods used for the proof of Theorems A and B lead also to the proof of the following result generalizing the characterization of diffeomorphisms satisfying Axiom A and the strong transversality condition given in [11]. In its statement we shall denote as usual $\Gamma^0(TM)$ the Banach space of continuous sections of TM and, for $f \in \mathrm{Diff}^1(M)$, $f_* = \Gamma^0(TM) \hookleftarrow$ will be the linear isomorphism defined as $f_*(\eta) = Tf \circ \eta \circ f^{-1}$.

Theorem C. *If $f_*: \Gamma^0(TM) \hookleftarrow$ has closed finite codimensional range then f satisfies Axiom A and the strong transversality condition.*

The plan of the proofs is the following: In Sect. 1 we shall consider a compact invariant set Λ of a diffeomorphism f and show that if Λ satisfies certain infinitesimal property then Λ has certain decomposition. This property when applied to $\Lambda = M$ will prove Theorem C. When applied to $\Lambda = \Gamma(f)$ will prove that $\Gamma(f)$ can be decomposed in the way described by Theorem B. Sections 2 and 3 will be devoted to show that the stability assumptions in Theorems A and B imply that the whole manifold M and $\Gamma(f_t)$ respectively satisfy the required infinitesimal property.

§1. Let $f \in \mathrm{Diff}^r(M)$. Denote T^*M the cotangent bundle of M and by T^*f: $T^*M \hookleftarrow$ the cotangent derivative of f i.e. the vector bundle isomorphism of T^*M covering f^{-1} and satisfying $\langle (T^*f)v, w\rangle = \langle v, (Tf)w\rangle$ for every $x \in M$, $w \in T_*M$, $v \in T^*_{f^{-1}(x)}M$. Let $\Lambda \subset M$ be a compact f-invariant set and let $\Gamma^b(\Lambda)$ be the Banach space of bounded sections of TM/Λ endowed with the norm $\|\eta\|_0 = \sup\{\|\eta(x)\| \mid x \in \Lambda\}$. Define $f_*: \Gamma^b(\Lambda) \hookleftarrow$ by $f_*(\eta) = Tf \circ \eta \circ f^{-1}$. Let $\Gamma^0(\Lambda)$ be the closed subspace of continuous sections of TM/Λ.

In this section we shall prove the following result:

Proposition 1.1. *If $\overline{(I-f_*)\,\Gamma^b(\Lambda)}$ contains a closed finite codimensional subspace of $\Gamma^0(\Lambda)$ then for every $x \in \Lambda \backslash \{x\}$ $\sup_n \|(T^*f)^n v\| = \infty$ for all $0 \neq v \in T^*_x M$. Moreover if Λ_2 is the set of non hyperbolic periodic points of f/Λ and $\Lambda_1 = \overline{\Lambda \backslash \Lambda_2}$ then Λ_2 is finite and $\Lambda_1 \cap \Lambda_2 = \emptyset$.*

Theorem C of the introduction is a corollary of 1.1. Under the hypothesis of Theorem C we can apply 1.1 to $\Lambda = M$. Since $x \in \overline{\Lambda \setminus \{x\}}$ for all $x \in M_x$ we obtain $\sup_n \|(T^* f)^n v\| = \infty$ for all $0 \neq v \in T^* M$. By [11] this implies that f satisfies Axiom A and the strong transversality condiction.

For the proof of 1.1 we shall need the following Lemmas:

Lemma 1.2. *If $x \in \overline{\Lambda \setminus \{x\}}$ and $\sup_n \|(T^* f)^n v\| < \infty$ for some $0 \neq v \in T_x^* M$ then there exists a neighborhood U of x and $n_0 > 0$ such that every $\bar{x} \in \Lambda \cap U$ satisfies $f^n(\bar{x}) = \bar{x}$ for some $0 < n \leqq n_0$.*

Proof. If $(I - f_*)\, \Gamma^b(\Lambda)$ contains a closed finite codimensional subspace of $\Gamma^0(\Lambda)$ then it is easy to see that $(I - f_*)\, \Gamma^b(\Lambda) \cap \Gamma^0(\Lambda)$ is closed and finite codimensional. Let $S = (I - f_*)^{-1}((I - f_*)\, \Gamma^b(\Lambda) \cap \Gamma^0(\Lambda))$. Then the adjoint linear map $(I - f_*)'$: $(\Gamma^0(\Lambda))' \to S'$, where S', $(\Gamma^0(\Lambda))'$ denote the dual spaces of S and $\Gamma^0(\Lambda)$, has closed range and a finite dimensional kernel L. We can write $(\Gamma^0(\Lambda))' = L \oplus P$ for some closed subspace P. Then $(I - f_*)'/P \colon P \to S'$ is an isomorphism into and so $\|(I - f_*)'\, \xi'\| \geqq k\, \|\xi'\|$ for some $k > 0$ and every $\xi' \in P$. Suppose that x, v are as in the statement of the Lemma. We claim that there exists $n > 0$ such that:

$$\sum_{|m| \leqq n} \|(T^* f)^m v\| \geqq \frac{k}{2} \left(\|(T^* f)^{n+1} v\| + \|(T^* f)^{-n} v\|\right). \tag{1}$$

If such an n doesn't exist we should have

$$\sum_{|m| \leqq n} \|(T^* f)^m v\| < \left(\frac{k}{2}\right) \left(\|T^* f)^{n+1} v\| + \|(T^* f)^{-n} v\|\right) \tag{2}$$

for arbitrary large values of n. In particular

$$\sum_{|m| \leqq n} \|(T^* f)^m v\| < \left(\frac{k}{2}\right) 2 \sup_n \|(T^* f)^n v\|.$$

Hence:

$$\sum_{-\infty}^{\infty} \|(T^* f)^n v\| < \infty$$

therefore $\lim_{n \to \pm\infty} \|(T^* f)^n v\| = 0$ and this together with (2) implies $\|(T^* f)^m v\| = 0$ for all m, hence $v = 0$. Then (1) is true for some $n > 0$. Since $x \in \overline{\Lambda \setminus \{x\}}$ if the Lemma is false we can take a sequence $x_m \in \Lambda \setminus \{x\}$, $m \in \mathbb{Z}^+$ satisfying $x_m \to x$ if $m \to +\infty$, and

$$f^j(x_{m''}) \neq f^i(x_{m'}) \tag{3}$$

for all $0 < m' \leqq m''$, $-n \leqq j < i \leqq n+1$. Take $v_m \in T_{x_m}^* M$ satisfying $\lim_{m \to \infty} v_m = v$. Define $\xi_m' \in (\Gamma^0(\Lambda))'$ by

$$\langle \xi_m', \eta \rangle = \sum_{|j| \leqq n} \langle (T^* f)^j v_m, \eta(f^{-j}(x_n)) \rangle.$$

Then:

$$\langle (I-f_*)'\,\xi'_m, \eta\rangle = \langle \xi'_m, (I-f_*)\,\eta\rangle = \langle (T^*f)^{-m}\,v_m, \eta(f^n(x_m))\rangle \\ -\langle (T^*f)^{(m+1)}\,v_m, \eta(f^{-(m+1)}(x_m))\rangle.$$

Using (3) it follows that

$$\|\xi'_m\| = \sum_{|j|\leqq n} \|(T^*f)^j\,v_m\|,$$
$$\|(I-f_*)\,\xi'_m\| = \|(T^*f)^{-m}\,v_m\| + \|(T^*f)^{(m+1)}\,v_m\|.$$

By (1) and the property $\lim_{m\to+\infty} v_m = v$:

$$\|(I-f_*)'\,\xi'_m\| \leqq \tfrac{2}{3}k\,\|\xi'_m\|$$

if m is large enough, say $m\geqq m_0$. If ξ' belong to the space spanned by the ξ'_ms with $m\geqq m_0$ it can be written as

$$\xi' = \sum_{j=1}^{r} \lambda_j\,\xi'_{m_j}$$

and then:

$$\|(I-f_*)'\,\xi'\| \leqq \sum_{j=1}^{r} |\lambda_j|\,\|(I-f_*)'\,\xi'_{m_j}\| \leqq \tfrac{2}{3}k \sum_{j=1}^{r} |\lambda_j|\,\|\xi'_{m_j}\|.$$

But using (3) it is easy to see that

$$\|\xi'\| = \sum_{j=1}^{r} |\lambda_j|\cdot\|\xi'_{m_j}\|$$

hence:

$$\|(I-f_*)'\,\xi'\| \leqq \tfrac{2}{3}k\,\|\xi'\|$$

for every ξ' in the space spanned by $\{\xi'_m\,|\,m\geqq m_0\}$. But this space has infinite dimension so it must have non trivial intersec intersection with P thus contradicting the property $\|(I-f_*)'\,\xi'\|\geqq k\,|\xi'\|$ for all $\xi'\in P$.

Lemma 1.3. *There exists $\varepsilon>0$ such that the set of periodic points $x\in\Lambda$ such that there exists $0\neq v\in T_x^*M$ satisfying $\|(T^*f)^n\,v-v\|\leqq\varepsilon\,\|v\|$, where n is the prime period of x, is finite.*

Proof. Suppose the Lemma false. Then there exist sequences of periodic points $x_n\in\Lambda$ and vectors $0\neq v_n\in T^*M$ such that if m_n is the prime period of x_n satisfy $\|(T^*f)^{m_n}\,v_n - v_n\| \leqq \dfrac{1}{n}\,\|v_n\|$ and $x_i\neq x_j$ if $i\neq j$. Let $\xi'_n\in(T^0(\Lambda))'$ defined by:

$$\langle \xi'_n, \eta\rangle = \sum_{j=0}^{m_n-1} \langle (T^*f)^j\,v_n, \eta(f^{-j}(x_n))\rangle. \tag{1}$$

Then:

$$\langle (I-f_*)\,\xi'_n, \eta\rangle = \langle (I-(T^*f)^{m_n})\, v_n, \eta(x)\rangle. \tag{2}$$

Since m_n is the prime period of x_n it follows from (1) that

$$\|\xi'_n\| = \sum_{j=0}^{m_n-1} \|(T^*f)^j\, v_n\| \geqq \|v_n\|.$$

By (2):

$$\|(I-f_*)\,\xi'_n\| = \|v_n - (T^*f)^{m_n}\, v_n\|$$

then:

$$\|(I-f_*)\,\xi'_n\| \leqq \frac{1}{n}\,\|v_n\| \leqq \frac{1}{n}\,\|\xi'\|. \tag{3}$$

Let $k>0$ be the same that in the proof of 1.2. Take $n_0 \geqq 2/k$. As in the previous proof from (3) follows that $\|(I-f_*)\,\xi'\| \leqq (k/2)\,\|\xi'\|$ for every ξ' in the space spanned by $\{\xi'_n \mid n \geqq n_0\}$. Again as in Lemma 1.2 this space must intersect P because it is infinite dimensional thus contradicting the fundamental property of k.

Lemma 1.4. *For every $N>0$ the set of fixed points of f^N in Λ is finite.*

Proof. If the Lemma is false we can take a sequence $\{x_n \mid n \in \mathbb{Z}^+\} \subset \Lambda$ of periodic points such that all of them have the same prime period n_1 and $x_{n'} \neq x_{n''}$ if $n' \neq n''$. We can assume that $x_n \to x \in \Lambda$ when $n \to +\infty$ and then $f^{n_1}(x) = x$. Since x is an accumulation point of fixed points of f^{n_1} it follows that $(Tf)^{n_1}/(T_x M)$ has an eigenvalue 1. Then $(T^*f)^{n_1}/(T^*_x M)$ has the same property. Let $0 \neq v \in T^*_x M$ with $(T^*f)^{n_1}\, v = v$. Take a sequence $v_n \in T^*_{x_n} M$ such that $\lim v_n = v$. Then $\|(T^*f)^{m_n}\, v_n - v_n\| / \|v_n\| \to 0$ when $n \to +\infty$ contradicting 1.3.

Now we can prove Proposition 1.1. First let us show that $\sup_n \|(T^*f)^n\, v\| = \infty$ if $0 \neq v \in T^* M$ and $x \in \overline{\Lambda \setminus \{x\}}$. If the property is false i.e. if $\sup_n \|(T^*f)^n\, v\| < \infty$ for some $0 \neq v \in T^*_x M$ with $x \in \overline{\Lambda \setminus \{x\}}$ then $x = \lim_{n\to\infty} x_n$ where $x_n \in \Lambda$ is periodic, $x_n \neq x$ for all n and the prime periods of the x'_ns are bounded (this follows from Lemma 1.2). So by 1.4 this sequence can contain only finitely many points thus implying $x_n = x$.

Now we shall prove that Λ_2 if finite. If it is not it contains a sequence x_n, $n \in \mathbb{Z}^+$ with infinitely many different points. Assume that $x_n \to x$ if $n \to +\infty$. Since each x_n is a non hyperbolic periodic point there exists $v_n \in T^*_{x_n} M$ with $\sup_m \|(T^*f)^m\, v_n\| = 1$ and $\|v_n\| = 1$. Suppose $v_n \to v \in T^*_x M$ when $n \to +\infty$. Then $\|v\| = 1$ and $\sup_m \|(T^*f)^m\, v\| = 1$. By Lemma 1.2 this implies that the periods of the x'_ns are bounded and then by Lemma 1.4 this means that there are only finitely many different x'_ns.

It remains to show $\Lambda_1 \cap \Lambda_2 = \emptyset$. Suppose $x \in \Lambda_1 \cap \Lambda_2$. By $x \in \Lambda_2$ there exists $0 \neq v \in T_x^* M$ with $\sup_n \|(T^*f)^n v\| < \infty$. Moreover $x \in \Lambda_2$ implies that $x \notin \overline{\Lambda \setminus \Lambda_2}$. But $x \in \Lambda_1 = \overline{\Lambda \setminus \Lambda_2}$. Hence $x \in \overline{\Lambda \setminus \{x\}}$ contradicting the existence of a $0 \neq v \in T_x^* M$ with $\sup_n \|(T^*f)^n v\| < \infty$.

Proposition 1.5. *If $f \in \mathrm{Diff}^r(M)$ and $(I - f_*)\Gamma^b(\Gamma(f))$ contains a closed finite codimensional subspace of $\Gamma^0(\Gamma(f))$ then $\Gamma(f)$ can be decomposed as the disjoint union of a finite set Γ_2 of non hyperbolic periodic points, an isolated AS set Γ_0 and a finite set $\gamma_1, \ldots, \gamma_m$ of non transversal orbits with $\alpha(\gamma_i) \cup \omega(\gamma_i) \subset \Gamma_0$. When $\Gamma(f)$ doesn't contain non transversal orbits $\Gamma(f/\Gamma_0) = \Gamma_0$.*

Proof. By 1.1 we can write $\Gamma(f) = \Gamma_1 \cup \Gamma_2$ where Γ_2 is a finite set of non hyperbolic periodic orbits, $\Gamma_1 = \overline{\Gamma(f) \setminus \Gamma_2}$ and $\Gamma_1 \cap \Gamma_2 = \emptyset$. Let Γ_1' be the set of $x \in \Gamma_1$ such that $\sup_n \|(T^*f)^n v\| < \infty$ for some $0 \neq v \in T_x^* M$. We claim that Γ_1' contains only finitely many orbits. Otherwise we could find a sequence $x_n \in \Gamma_1'$, $n \in \mathbb{Z}^+$ with $x_{n'} \neq f^m(x_{n''})$ for all $m \in \mathbb{Z}$ $n' \neq n''$ and a sequence $v_n \in T_{x_n}^* M$ with $\|v_n\| = 1$ and $\sup_m \|(T^*f)^m v_n\| = K_n < \infty$. Let $m_n \in \mathbb{Z}$ such that $\|(T^*f)^{m_n} v_n\| \geqq K_n/2$. Then if $w_n = (T^*f)^{m_n} v_n$ we have $\|w_n\| \geqq K_n/2 \geqq 1/2$ and $\sup_m \|(T^*f)^m w_n\| \leqq 2$. Suppose that $\lim_{n \to +\infty} f^{m_n}(x_n) = x$, $\lim_{n \to +\infty} w_n = w \in T_x^* M$. Then $\|w\| \geqq 1/2$ and $\sup_m \|(T^*f)^m w\| \leqq 2$. By

Proposition 1.1. $x \notin \overline{\Gamma(f) \setminus \{x\}}$. *This means that $f^{m_n}(x_n) = x$ for every large n contradicting the fact that $x_{n'} \neq f^m(x_{n''})$ for all $m \in \mathbb{Z}$ if $n' \neq n''$. Later we shall show that the orbits in Γ_1' are precisely the non transversal orbits. Before that we shall show that Γ_0 defined by $\Gamma_0 = \Gamma_1 \setminus \Gamma_1'$ is an isolated AS set.*

This set is closed because if $x_n \in \Gamma_0$ and $\lim_{n \to +\infty} x_n = x$ then either $x_n = x$ for some n (and then $x = x_n \in \Gamma_0$) or $x \in \overline{\Gamma(f) \setminus \{x\}}$. By 1.1 $x \notin \Gamma_2 \cup \Gamma_1'$, then $x \in \overline{\Gamma(f) \setminus (\Gamma_2 \cup \Gamma_1')} = \overline{(\Gamma(f) \setminus \Gamma_2) \setminus \Gamma_1'} \subset \Gamma_1 \setminus \Gamma_1' = \Gamma_0$. Every $0 \neq v \in T^* M/\Gamma_0$ satisfies $\sup_n \|(T^*f)^n v\| = \infty$. By [17] T^*f is a hyperbolic isomorphism $T^* M/\Gamma(f/\Gamma_0)$. Then so is Tf. This means that $\Gamma(f/\Gamma_0)$ is a hyperbolic set for f. Let us show that it has local product structure [7]. Denote by $W_\varepsilon^s(x)$, $W_\varepsilon^u(x)$, $x \in \Gamma(f/\Gamma_0)$, the local stable and unstable manifolds [7]. Suppose that $z \in W_\varepsilon^s(x) \cap W_\varepsilon^u(y)$. If ε is small enough this means $W_{2\varepsilon}^s(y) \cap W_{2\varepsilon}^u(x) \neq \emptyset$. Let $w \in W_{2\varepsilon}^s(y) \cap W_{2\varepsilon}^u(x)$. Since $z \in W_\varepsilon^s(x)$, given any $\delta > 0$ we can find a δ-orbit joining z with $f^{N_1}(x)$ if N_1 is large enough i.e. a set $z = x_0$, $x_1, \ldots, x_l = f^{N_1}(x)$ satisfying $d(f(x_j), x_{j+1}) \leqq \delta$ for all $0 \leqq j \leqq l$. In a similar way we find δ-orbits joining z and $f^{-N_2}(y)$, w and $f^{-N_3}(x)$ and w and $f^{N_4}(y)$, with $N_2 > 0$, $N_3 > 0$, $N_4 > 0$ large enough. Moreover since $y, x \in \Gamma(f/\Gamma_0)$ we can find δ-orbits joining $f^{N_1}(x)$ with $f^{-N_3}(x)$ and $f^{-N_2}(y)$ with $f^{N_4}(y)$. The union of these δ-orbits gives a 2δ-orbit joining z with itself. Since this can be done for every $\delta > 0$ we have proved $z \in \Gamma(f)$. Now take a neighborhood V of Γ_0 such that $V \cap \Gamma_2 = \emptyset$ (this can be done because $\Gamma_0 \cap \Gamma_2 \subset \Gamma_1 \cap \Gamma_2 = \emptyset$) and not containing any orbit of Γ_1' (this is possible because Γ_1' contains only finitely many orbits). If ε is small enough the whole orbit of z is contained in V. Therefore $z \notin \Gamma_2 \cup \Gamma_1'$. Hence $z \in \Gamma(f) \cap \Gamma_0 = \Gamma_0$. In a similar way it follows that $w \in \Gamma_0$. But then observe that all the δ-orbits constructed in the proof of $z \in \Gamma(f)$ can be obtained contained in Γ_0 thus showing that $z \in \Gamma(f/\Gamma_0)$. The local product structure of $\Gamma(f/\Gamma_0)$ and the fact that $f/\Gamma(f/\Gamma_0)$ is chain recurrent ([4]) imply ([1, 4]) that the periodic points of f are dense in $\Gamma(f/\Gamma_0)$.

Now we shall prove the transversality condition i.e. that if $x \in W^s(x_1) \cap W^u(x_2) \cap \Gamma_0$ with $x_1, x_2 \in \Gamma(f/\Gamma_0)$ then $T_x W^u(x_1) + T_x W_\varepsilon^s(x_2) = T_x M$. Define the polar spaces:

$$(T_x W_\varepsilon^s(x_1))^0 = \{v \in T_x^* M \mid \langle v, w\rangle = 0 \text{ for all } w \in T_x W_\varepsilon^s(x_1)\},$$

$$(T_x W^u(x_2))^0 = \{v \in T_x^* M \mid \langle v, u\rangle = 0 \text{ for all } u \in T_x W^u(x_2)\}.$$

As in [11] we can prove that $v \in (T_x W^s(x_1))^0$ implies $\sup\limits_{n \leqq 0} \|(T^* f)^n v\| < \infty$ and $v \in (T_x W^u(x_2))^0$ implies $\sup\limits_{n \geqq 0} \|(T^* f)^n v\| < \infty$. Hence if $T_x W^s(x_1) + T_x W^u(x_2) \neq T_x M$ we should have

$$(T_x W^s(x_1))^0 \cap (T_x W^u(x_2))^0 \supset (T_x W^s(x_1) + T_x W^u(x_2))^0 \neq \{0\}.$$

Therefore there exists

$$0 \neq v \in (T_x W^s(x_1))^0 \cap (T_x W^u(x_2))^0$$

and then $\sup\limits_n \|(T^* f)^n v\| < \infty$ contradicting $x \in \Gamma_0$.

To show that Γ_0 is isolated we first introduce the following equivalence relation in $\Gamma(f)$: $x_1 \sim x_2$ if for all $\varepsilon > 0$ there exist $x_1 = y_0, y_1, \ldots, y_n = x_2$ satisfying $d(f(y_j), y_{j+1}) \leqq \varepsilon$ for all $0 \leqq j < n$. Let $\{F_\alpha \mid \alpha \in \mathscr{A}\}$ be the decomposition of $\Gamma(f)$ in equivalence classes. Let $\Gamma(f/\Gamma_0) = \Sigma_1 \cup \ldots \cup \Sigma_m$ be a decomposition of $\Gamma(f/\Gamma_0)$ in compact invariant disjoint sets such that f/Σ_i is transitive for all $1 \leqq i \leqq m$ (such a decomposition exists because the hyperbolicity and local product structure of $\Gamma(f/\Gamma_0)$). If $F_\alpha \cap \Sigma_i \neq \emptyset$ then $\Sigma_i \subset F_\alpha$ because f/Σ_i is transitive. The same property is obviously true replacing Σ_i by any periodic orbit. Therefore if we show that every class F_α intersects a Σ_i or a non hyperbolic periodic orbit it will follow that there are only finitely many equivalence classes. To prove that property let F_α be an equivalence class and $x \in F_\alpha$. Then $\omega(x) \subset F_\alpha$. Hence either x is periodic (and then F_α intersects $\Gamma(f/\Gamma_0)$ or Γ_2) or $y \in \overline{\Gamma(f) \setminus \{y\}}$ for every $y \in \omega(x)$. By 1.1 this means that $y \notin \Gamma_1'$. Hence $y \in \Gamma_0 \cup \Gamma_2$. If $y \in \Gamma_2$ we are done. If $y \in \Gamma_0$ it follows that $F_\alpha \cap \Gamma_0 \neq \emptyset$. Since F_α is closed and invariant this implies that $F_\alpha \cap \Gamma(f/\Gamma_0) \neq \emptyset$ completing the proof. Let $F_1, \ldots, F_r$ be the equivalence classes. Let U_i be a neighborhood of $F_i \cap \Gamma_0$, $i = 1, \ldots, k$ satisfying:

1) $(\bigcup\limits_i U_i) \cap \Gamma_2 = \emptyset, \quad \bigcup\limits_{j=1}^{k} (F_j \cap \Gamma_0) = \Gamma_0,$

2) $U_i \cap (U_j \cup f(U_j)) = \emptyset \quad$ for all $i \neq j$,

3) U_i doesn't contain orbits contained in Γ_1'.

We claim that if $U = \bigcup\limits_i U_i$ then

$$\bigcap_n f^n(U) = \Gamma_0.$$

By (2)

$$\bigcap_n f^n(U) = \bigcup_i \Big(\bigcap_n f^n(U_i)\Big)$$

so it is sufficient to prove:

$$\bigcap_n f^n(U_i)=\Gamma_0\cap F_i.$$

If $x\in\bigcap_n f^n(U_i)$ then the whole orbit of x is contained in U_i, hence $x\notin\Gamma_1'$. By (1) $x\notin\Gamma_2$. So to prove that $x\in\Gamma_0\cap F_i$ it is sufficient to show that $x\in\Gamma$ because that will imply $x\in\Gamma\setminus(\Gamma_1'\cup\Gamma_2)=\Gamma_0$ and since $x\in U_i$ we have

$$x\in U_i\cap\Gamma_0=U_i\cap\left(\bigcup_{j=1}^{k}(F_j\cap\Gamma_0)\right)\subset U_i\cap((\bigcup_{j\neq i}U_j)\cup(F_i\cap\Gamma_0))=U_i\cap(F_i\cap\Gamma_0)=F_i\cap\Gamma_0.$$

To show $x\in\Gamma$ observe that $\alpha(x)\cup\omega(x)\subset U_i$. Hence $\alpha(x)\cup\omega(x)\subset\Gamma_0\cap U_i$. By the same argument that before $\Gamma_0\cap U_i\subset F_i\cap\Gamma_0$. Then $\alpha(x)\subset F_i$, $\omega(x)\subset F_i$. Given $\varepsilon>0$ take $n>0$, $x_1\in\omega(x)$, $x_2\in\alpha(x)$ satisfying $d(f^n(x),x_1)\leqq\varepsilon/2$, $d(f^{-n}(x),x_2)\leqq\varepsilon/2$. Since x_1,x_2 belong to the same equivalence class we can find a set $x_1=y_0,\dots,y_m=x_2$ satisfying $d(f(y_j),y_{j+1})\leqq\varepsilon/2$ for every $0\leqq j\leqq m$. Hence the set $x,f(x),\dots,f^{n_1}(x)$, $x_1=y_0,y_1,\dots,y_m=x_2,f^{-n_2}(x),\dots,x$ is an ε-chain thus proving $x\in\Gamma$.

Finally let us show that any orbit $\gamma\subset\Gamma_1'$ is non transversal and $\alpha(\gamma)\cup\omega(\gamma)\subset\Gamma_0$. The second property follows from the fact that $\alpha(\gamma)\cup\omega(\gamma)\in\Gamma_1$ (because $\gamma\subset\Gamma_1$ and Γ_1 is closed). Hence $(\alpha(\gamma)\cup\omega(\gamma))\not\subset\Gamma_0$ would imply $(\alpha(\gamma)\cup\omega(\gamma))\cap\Gamma_1'\neq\emptyset$. But since γ is not periodic this means that Γ_1' contains points $x\in\alpha(\gamma)\cup\Gamma_1'$ that satisfy $x\in\overline{\Gamma(f)\setminus\{x\}}$ contradicting 1.1 and the definition of Γ_1'. To show that γ is non transversal take $x\in\gamma$ and define

$$E^s(x)=\{v\in T_xM\mid \lim_{n\to+\infty}\|(Tf)^n v\|=0\},$$
$$E^u(x)=\{v\in T_xM\mid \lim_{n\to+\infty}\|(Tf)^{-n} v\|=0\},$$
$$E_*^s(x)=\{v\in T_x^*M\mid \sup_{n\geqq 0}\|(T^*f)^n v\|<\infty\},$$
$$E_*^u(x)=\{v\in T_x^*M\mid \sup_{n\geqq 0}\|(T^*f)^{-n} v\|<\infty\}.$$

Moreover let $(E^s(x))^0\subset T_x^*M$, $(E^u(x))^0\subset T_x^*M$ be the polar spaces of $E^s(x)$, $E^u(x)$. We claim that

$$(E^s(x))^0\supset E^u_*(x), \tag{1}$$
$$(E^u(x))^0\supset E^s_*(x). \tag{2}$$

If $v\notin(E^s(x))^0$ there exists $w\in E^s(x)$ such that $\langle v,w\rangle\neq 0$. Then

$$\|(T^*f)^{-n}v\|\,\|(Tf)^n w\|\geqq|\langle(T^*f)^{-n}v,(Tf)^n w\rangle|=|\langle v,w\rangle|.$$

Then:

$$\lim_{n\to+\infty}\|(T^*f)^{-n}v\|=\lim_{n\to+\infty}\frac{|\langle v,w\rangle|}{\|(Tf)^n w\|}=0$$

implying that $v \notin E^u_*(x)$. In a similar way we prove (2). But $x \in \Gamma_1'$ implies $E^s_*(x) \cap E^u_*(x) \neq \{0\}$. Hence

$$\{0\} \subsetneqq E^u_*(x) \cap E^s_*(x) \subset (E^s(x))^0 \cap (E^u(x))^0 = (E^s(x)+E^u(x))^0$$

then $E^s(x)+E^u(x) \neq T_* M$. Moreover it is easy to see that $\alpha(\gamma) \subset \Gamma_0$, $\omega(\gamma) \subset \Gamma_0$ imply $\alpha(\gamma) \subset \Gamma(f/\Gamma_0)$, $\omega(\gamma) \subset \Gamma(f/\Gamma_0)$. Since $\Gamma(f/\Gamma_0)$ is a hyperbolic set with local product structure there exist [1, 4] $y_1 \in \Gamma(f/\Gamma_0)$ $y_2 \in \Gamma(f/\Gamma_0)$ such that

$$\lim_{n \to +\infty} d(f^n(x), f^n(y_1)) = \lim_{n \to -\infty} d(f^n(x), f^n(y_2)) = 0.$$

It is easy to see that $T_x W^s(y_1) = E^s(x)$ and $T_x W^u(y_2) = E^u(x)$. Hence $T_x M \neq T_x W^s(y_1) + T_x W^u(y_2)$. It remains to show that when f doesn't contain non transversal orbits then $\Gamma(f/\Gamma_0) = \Gamma_0$. Suppose that $x \in \Gamma_0 \setminus \Gamma(f/\Gamma_0)$. For every $n \in \mathbb{Z}^+$ take a set $x = x_0^{(n)}, \ldots, x_{r_n}^{(n)} = x$ satisfying $d(f(x_j^{(n)}),\ x_{j+1}^{(n)}) \leqq 1/n$ for all $0 \leqq j < r_n$. To show that $x \in \Gamma(f/\Gamma_0)$ it is sufficient to prove that if $\delta_n = \sup\{d(x_j^{(n)}, \Gamma_0) \mid 0 \leqq j < r_n\}$ then $\lim_{n \to +\infty} \delta_n = 0$. If $\liminf_{n \to +\infty} \delta_n = \delta > 0$ there exists a sequence

$$\{n_k \mid k \in \mathbb{Z}^+\} \subset \mathbb{Z}^+ \quad \text{with} \quad \lim_{k \to +\infty} n_k = \infty \quad \text{and} \quad x_k \in \{x_j^{(n)} \mid 0 \leqq j \leqq r_n\}$$

such that $d(x_k, \Gamma_0) \geqq \delta/2$. We can suppose that x_k converges to some $\bar{x} \in M$ when $k \to +\infty$. Then $\bar{x} \in \Gamma$ and since $\bar{x} \notin \Gamma_0$ and $\Gamma_1' = \emptyset$ we must have $\bar{x} \in \Gamma_2$. Take open neighborhoods U_0, U_2 of Γ_0, Γ_2. Let $\varepsilon = \inf\{d(p \cdot q) \mid p \in U_2 \cup f(U_2),\ q \in U_0\}$. We can suppose $\varepsilon > 0$. If $n_k \geqq (\varepsilon/2)^{-1}$ let $j_k = \sup\{j \mid x_j^{(n_k)} \in U_2\}$ (this set is non empty for large values of k because $\bar{x} \in \Gamma_2$). Since $x_{r_n} = x \in U_0$ we must have $j_k < r_n$. Then $d(f(x_{j_k}), x_{j_k+1}) \leqq 1/n_k \leqq \varepsilon/2$. Since $f(x_{j_k}) \in f(U_2)$ and by the definition of ε we have $x_{j_k+1} \notin U_0$. By definition of j_k we have $x_{j_k+1} \notin U_2$. Then $x_{j_k+1} \notin U_0 \cup U_2$. Suppose that when $k \to +\infty$ the sequence x_{j_k} converges to some $\hat{x}$. Then $\hat{x} \in \Gamma$. But $\Gamma = \Gamma_0 \cup \Gamma_2 \subset U_0 \cup U_2$ contradicting $\hat{x} \notin U_0 \cup U_2$.

§2. In this section we shall consider a diffeomorphism $f \in \mathrm{Diff}^r(M)$ with a compact invariant set $\Lambda \subset M$ satisfying the following property: there exist a subset $\Sigma \subset \mathrm{Diff}^r(M)$ containing f, a constant $k > 0$ and a neighborhood $\mathcal{U}$ of f in $\mathrm{Diff}^r(M)$ such that if $g \in \mathcal{U} \cap \Sigma$ then there exists a continuous map $h\colon \Lambda \to M$ such that $gh(x) = hf(x)$ for all $x \in \Lambda$ and $d_0(h, I) \leqq k d_0(f, g)$. Define the tangent cone $C_f(\Sigma)$ of Σ at f as the set of all $\xi \in \Gamma^0(TM)$ such that there exists a sequence ξ_n of C^r sections of TM and a sequence $\lambda_n \in \mathbb{R}$, $n \in \mathbb{Z}^+$ such that $\exp \xi_n \circ f \in \Sigma$, $\exp \xi_n \circ f \to f$ when $n \to +\infty$ in $\mathrm{Diff}^r(M)$ and $\lim_{n \to +\infty} \|\lambda_n \xi_n - \xi\|_0 = 0$ (recall that $\|\cdot\|_0$ denotes the C^0 norm in $\Gamma^0(TM)$).

Proposition 2.1. *There exists $K > 0$ such that for every $\xi \in C_f(\Sigma)$ there exists $\eta \in \Gamma^b(\Lambda)$ satisfying $\|\eta\|_0 \leqq K \|\xi/\Lambda\|_0$ and $(I - f_*)\eta = \xi/\Lambda$.*

Proof. Let $\Gamma^r(TM)$ be the space of C^r sections of TM with the usual C^r norm $\|\cdot\|_r$. We shall use the following property proved in [5]: If $\mathcal{U}$ is small enough there exists $\delta_0>0$ and a function $P\colon B\to\Gamma^0(\Lambda)$ where

$$B=\{(\xi,\eta)\mid \xi\in\Gamma^r(TM),\ \eta\in\Gamma^0(\Lambda),\ \|\xi\|_r\leqq\delta_0,\ \|\eta\|_0\leqq\delta_0\}$$

such that if $\xi\in\Gamma^r(TM)$, $\|\xi\|_r\leqq\delta_0$, $g=\exp\xi\circ f\in\mathscr{U}$ and there exists $\eta\in\Gamma^0(\Lambda)$ such that $((\exp\eta)\circ f)(x)=(g\circ(\exp\eta))(x)$ for every $x\in\Lambda$ then:

$$(I-f_*)\,\eta=\xi/\Lambda+P(\xi,\eta).$$

Moreover there exists $C>0$ such that for all $\varepsilon>0$ there exists $0<\delta\leqq\delta_0$ such that

$$\|P(\xi,\eta)\|_0\leqq(C\,\|\xi\|_r+\varepsilon)\,\|\eta\|_0 \tag{1}$$

for all $(\xi,\eta)\in B$ with $\|\xi\|_r\leqq\delta_0$.

To prove 2.1 let $\xi\in C_f(\Sigma)$ with $\xi=\lim\limits_{n\to+\infty}\lambda_n\xi_n$ in $\Gamma^0(TM)$ with $\lambda_n\in\mathbb{R}$ and $\xi_n\in\Gamma^r(TM)$ satisfying $\exp\xi\circ f\in\Sigma$ and $\lim\limits_{n\to+\infty}\|\xi_n\|_r=0$. Let $g_n=\exp\xi_n\circ f$. We can suppose $\|\xi_n\|_r\leqq\delta_0$ and $g_n\in\mathscr{U}$. Hence there exists a continuous map $h_n\colon\Lambda\to M$ satisfying $d_0(h_n,I)\leqq k\,d_0(g_n,f)$ and $g_n h_n(x)=h_n f(x)$ for all $x\in\Lambda$. Write $h_n=\exp\eta_n$, $\eta_n\in\Gamma^0(\Lambda)$. We can suppose that for some $\bar{k}>0$ we have $\|\eta_n\|_0\leqq\bar{k}\,\|\xi\|_0$ for all n. Then

$$(I-f_*)\,\eta_n=\xi_n/\Lambda+P(\xi_n,\eta_n)$$

hence:

$$(I-f_*)\,\lambda_n\eta_n=\lambda_n\xi_n/\Lambda+\lambda_n P(\xi_n,\eta_n)$$

and by (1):

$$\|\lambda_n P(\xi_n,\eta_n)\|_0\leqq|\lambda_n|\,(C\,\|\xi_n\|_r+\varepsilon_n)\,\|\eta_n\|_0$$

with $\varepsilon_n\to 0$. Then if n is large enough to have $\|\lambda_n\xi_n\|_0\leqq 2\,\|\xi\|_0$ we obtain:

$$\begin{aligned}\|\lambda_n P(\xi_n,\eta_n)\|_0&\leqq(C\,\|\xi_n\|_r+\varepsilon_n)\,|\lambda_n|\,\bar{k}\,\|\xi_n\|_0\\&\leqq 2(C\,\|\xi_n\|_r+\varepsilon_n)\,\bar{k}\,\|\xi\|_0.\end{aligned} \tag{2}$$

Moreover the sequence $\lambda_n\eta_n$ is bounded because $\|\lambda_n\eta_n\|_0\leqq\bar{k}\,\|\lambda_n\xi_n\|_0\leqq 2\bar{k}\,\|\xi\|_0$ if n is large. Let N be the space of vectors $v\in TM/\Lambda$ such that $\|v\|\leqq 2\bar{k}\,\|\xi_0\|$. N is compact and then, by Tijonov's theorem, the set of maps of Λ in N endowed with the pointwise convergence topology is a compact space (because it coincides with the product space $N^{\wedge}$). Therefore the sequence ξ_n, that it is contained in this space, has a subnet $\{\gamma_\alpha\mid\alpha\in\mathscr{A}\}$ where α is a directed set, that is pointwise convergent to $\gamma\colon\Lambda\to N$, i.e., $\gamma(x)=\lim\limits_{\alpha}\gamma_\alpha(x)$. It is easy to see that γ is a section of TM/Λ and since $\gamma(\Lambda)\subset N$ we have $\|\gamma(x)\|\leqq 2\bar{k}\,\|\xi\|_0$ that means $\|\gamma\|_0\leqq 2\bar{k}\,\|\xi\|_0$. Let $\varphi\colon\mathscr{A}\to\mathbb{Z}^+$ be the map given by the property of $\{\gamma_\alpha\mid\alpha\in\mathscr{A}\}$ be a subnet of $\{\eta_n\mid n\in\mathbb{Z}^+\}$. Then

$$(I-f_*)\,\gamma_\alpha=\lambda_{\varphi(\alpha)}\,\xi_{\varphi(\alpha)}/\Lambda+\lambda_{\varphi(\alpha)}\,P(\xi_{\varphi(\alpha)},\gamma_\alpha).$$

Taking limit on $\alpha \in \mathscr{A}$ at each $x \in M$ and using (2) we conclude

$$(I - f_*)\gamma = \xi/\Lambda.$$

Putting $K = 2\bar{k}$ we also have $\|\gamma\|_0 \leqq K \|\xi\|_0$ as desired.

§3. In this section we shall prove the following Proposition:

Proposition 3.1. *Let $f \in F^{k,r}(V, M)$, $t \in V$ and Λ a compact invariant set of f_t satisfying the following property: there exist $k_1 > 0$ and a neighborhood $\mathscr{U}$ of f in $F^{k,r}(V, M)$ such that for every $g \in \mathscr{U}$ there exists $\bar{t} \in V$ and a homeomorphism $h: \Lambda \to M$ satisfying:*

1) $g_{\bar{t}}\, h(x) = h_t f(x)$ *for all* $x \in \Lambda$,
2) $d_0(h, I) \leqq k_1\, d_0(g, f)$,
3) $d(\bar{t}, t) \leqq k_1\, d_0(g, f)$

then $(I - f_{t_*})\, \Gamma^b(\Lambda)$ contains a closed finite codimensional subspace of $\Gamma^0(\Lambda)$.

If f is absolutely Γ-stable at t we can apply 3.1 to $\Lambda = \Gamma(f_t)$. Together with Proposition 2.5 this proves Theorem B of the introduction. If f is absolutely stable a t we can apply 3.1 to $\Lambda = \Gamma(f_t)$ thus obtaining that $(I - f_*)\, \Gamma^b(TM)$ contains a closed finite codimensional subspace of $\Gamma^0(TM)$. Hence $(I - f_*)\, \Gamma^0(TM)$ will have the same property. As we observed after the statement of Proposition 1.1 it is easy to show that then $(I - f_*)\, \Gamma^0(TM)$ is a closed finite codimensional subspace of $\Gamma^0(TM)$. By Theorem C this implies that f satisfies Axiom A and the strong transversality condition.

To prove Proposition 3 we start considering the case when $f: V \to \mathrm{Diff}^r(M)$ is an immersion and then we shall show (using a method suggested by F. Takens) that if the Proposition holds for immersions then it is true in the general case. We shall prove first that there exists a set $\Sigma \subset \mathrm{Diff}^r(M)$ containing f_t such that Σ, Λ and f_t satisfy the hypothesis of Proposition 2.1. Let $A = T_{f_t} f(V)$ and take a closed subspace $B \subset \Gamma^r(TM)$ such that $\Gamma^r(TM) = A \oplus B$. Let $\mathscr{U}_0$ be a neighborhood of 0 in B such that $\mu \in B$ implies that $f^{(\mu)} = \exp \mu \circ f: V \to \mathrm{Diff}^r(M)$ belongs to $\mathscr{U}$. From the hypothesis on f it follows the existence of a constant $k_2 > 0$ and a map $\varphi: \mathscr{U}_0 \to V$ such that if $\mu \in \mathscr{U}_0$ then there exists a continuous $h_\mu: \Lambda \to M$ satisfying:

1′) $f^{(\mu)}_{\varphi(\mu)}\, h_\mu(x) = h_\mu f_t(x)$ for all $x \in \Lambda$,
2′) $d_0(h_\mu, I) \leqq k_2 \|\mu\|_0$,
3′) $d(\varphi(\mu), t) \leqq k_2 \|\mu\|_0$.

Let $\Sigma = \{f^{(\mu)}_{\varphi(\mu)} \mid \mu \in \mathscr{U}_0\}$. From (1′), (2′), (3′) and Proposition 2.1 follows that $(I - f_*)\, \Gamma^b(\Lambda)$ contains every ξ/Λ with $\xi \in C_{f_t}(\Sigma)$. Moreover there exists $K > 0$ such that every ξ/Λ with $\xi \in C_{f_t}(\Sigma)$ has a preimage η with $\|\eta\|_0 \leqq K \|\xi/\Lambda\|_0$. Now we shall prove some properties of $C_{f_t}(\Sigma)$.

Lemma. *There exists $k_3 > 0$ such that for every $v \in B$ we can find $w \in A$ satisfying $\|w\|_0 \leqq k_3 \|v\|_0$ and $v + w \in C_{f_t}(\Sigma)$.*

Proof. Let $\varphi_A: \mathscr{U}_0 \to A$, $\varphi_B: \mathscr{U}_0 \to B$ satisfying

$$\exp(\varphi_A(\mu) + \varphi_B(\mu)) \circ f_t = f^{(\mu)}_{\varphi(\mu)}.$$

From (3′) follows that there exists $k_3 > 0$ satisfying

$$\|\varphi_A(\mu)\|_0 \leqq k_3 \|\mu\|_0.$$

Moreover it is not difficult to prove that

$$\lim_{\|\mu\|_0 \to 0} \frac{\|\varphi_B(\mu) - \mu\|_0}{\|\mu\|_0} = 0.$$

Hence if $v \in B$ we have:

$$\lim_{\lambda \to 0} \left\| \frac{\varphi_B(\lambda v)}{\lambda} - v \right\|_0 = 0.$$

Moreover $\limsup_{\lambda \to 0} \|\varphi_A(\lambda v)/\lambda\|_0 < \infty$ because $\|\varphi_A(\lambda v)/\lambda\|_0 \leqq k_3 |\lambda| \, \|v\|_0/|\lambda| = k_3 \|v\|_0$. Hence we can take a sequence $\lambda_n \to 0$ such that $\varphi_A(\lambda_n v)/\lambda$ converges when $n \to +\infty$ to some $w \in A$ (recall that $\dim A < \infty$). Therefore the sequence $(\varphi_A(\lambda_n v) + \varphi_B(\lambda_n v))/\lambda_n$ converges in $\Gamma^0(TM)$ to $v + w$ and

$$\|w\|_0 \leqq \limsup_{n \to +\infty} \|\varphi_A(\lambda v)/\lambda\|_0 \leqq k_3 \|v\|_0$$

thus proving the Lemma.

Let $\bar{B}$ the closure of B in $\Gamma^0(TM)$ and $A_0 \subset A$ a subspace such that $A_0 \oplus (\bar{B} \cap A) = A$. Let us prove that $A_0 \oplus \bar{B} = \Gamma^0(TM)$. If $\xi \in \Gamma^0(TM)$ we can write $\xi = \lim_{n \to +\infty} \xi_n$ in $\Gamma^0(TM)$ with $\xi_n \in \Gamma^r(TM)$ and $\xi_n = \alpha_n + \beta_n$, $\alpha_n \in A_0$, $\beta_n \in \bar{B}$. If the sequence $\|\alpha_n\|_0$ is unbounded we can suppose $\lim_{n \to +\infty} \|\alpha_n\|_0 = \infty$ and that $\alpha_n / \|\alpha_n\|_0$ converges in $\Gamma^0(TM)$ to some $\alpha \in A_0$. Then $\xi_n / \|\alpha_n\|_0 = \alpha_n / \|\alpha_n\|_0 + \beta_n / \|\alpha_n\|_0$. Since $\alpha_n / \|\alpha_n\|_0$ converges so does $\beta_n / \|\alpha_n\|_0$. Hence:

$$0 = \alpha + \lim_{n \to +\infty} \frac{\beta_n}{\|\alpha_n\|_0}.$$

Since $\|\alpha\| = 1$ this contradicts $A_0 \cap \bar{B} \subset A_0 \cap (\bar{B} \cap A) = \{0\}$. Then $\sup_n \|\alpha_n\|_0 < \infty$ and we can assume (because A_0 is finite dimensional) that α_n converges. Since ξ_n is convergent and $\beta_n = \xi_n - \alpha_n$ then β_n is convergent and $\xi = \lim_{n \to +\infty} \alpha_n + \lim_{n \to +\infty} \beta_n$ proving $A_0 \oplus \bar{B} = \Gamma^0(TM)$. Now observe that for every $v \in B$ there exists $w \in A$ satisfying $w + v \in C_{f_t}(\Sigma)$ and $\|w\|_0 \leqq k_3 \|v\|_0$. To prove this take a sequence v_n, $n \in \mathbb{Z}^+$ contained in B converging to v and for each $n \in \mathbb{Z}^+$ a $w_n \in A$ satisfying $\|w_n\|_0 \leqq k_3 \|v_n\|_0$ and $w_n + v_n \in C_{f_t}(\Sigma)$. The sequence w_n is bounded and contained in the finite dimensional space A so we can suppose that converges to some $w \in A$. Hence $v + w \in C_{f_t}(\Sigma)$ (because $v_n + w_n \in C_{f_t}(\Sigma)$ and $C_{f_t}(\Sigma)$ is closed) and

$$\|w\|_0 \leqq \limsup_{n \to +\infty} \|w_n\|_0 \leqq \limsup_{n \to +\infty} k_3 \|v_n\|_0 = k_3 \|v\|_0.$$

To prove that $(I-f_{t_*})\,\Gamma^b(\Lambda)\cap\Gamma^0(\Lambda)$ is finite codimensional take an extension operator $\Gamma^0(\Lambda)\ni\xi\to\hat{\xi}\in\Gamma^0(TM)$ such that $\|\hat{\xi}\|\leqq 2\,\|\xi\|$ for all $\xi\in\Gamma^0(\Lambda)$. If $\xi\in\Gamma^0(\Lambda)$ write:

$$\hat{\xi}=\pi_B\,\hat{\xi}+\pi_{A_0}\hat{\xi}$$

where $\pi_{A_0}\colon\Gamma^0(TM)\to A_0$, $\pi_B\colon\Gamma^0(TM)\to\bar{B}$ are the projections associated to the splitting $\bar{B}\oplus A_0=\Gamma^0(TM)$. Take $w\in A$ with $\|w\|_0\leqq k_3\,\|\pi_B\hat{\xi}\|_0$ such that:

$$w+\pi_B\,\hat{\xi}\in C_{f_t}(\Sigma).$$

Then:

$$\begin{aligned}\xi+w/\Lambda-\pi_{A_0}\hat{\xi}/\Lambda&=(w+\hat{\xi}-\pi_{A_0}\hat{\xi})/\Lambda=(\pi_B\,\hat{\xi}+w)\in C_{f_t}(\Sigma)/\Lambda\\&\subset(I-f_{t_*})\,\Gamma^b(\Lambda)\cap\Gamma^0(\Lambda)\end{aligned}$$

since $+w/\Lambda-\pi_{A_0}\hat{\xi}/\Lambda\in A/\Lambda$ we conclude:

$$A/\Lambda+(I-f_{t_*})\,\Gamma^b(\Lambda)\cap\Gamma^0(\Lambda)=\Gamma^0(\Lambda).$$

To show that $(I-f_{t_*})\,\Gamma^b(\Lambda)\cap F^0(\Lambda)$ is closed it is sufficient to show that there exists $C>0$ such that every $\xi\in(I-f_{t_*})\,\Gamma^b(\Lambda)\cap\Gamma^0(\Lambda)$ has a preimage η with $\|\eta\|_0\leqq C\,\|\xi\|_0$. If $\xi\in(I-f_{t_*})\,\Gamma^b(\Lambda)\cap\Gamma^0(\Lambda)$ take $w\in A$ with $\|w\|_0\leqq k_3\,\|\pi_B\,\xi\|_0$ and $w+\pi_B\,\xi\in C_{f_t}(\Sigma)$. Then $\hat{\xi}=\pi_{A_0}\hat{\xi}+\pi_B\,\hat{\xi}=(\pi_{A_0}\hat{\xi}-w)+(\pi_B\,\hat{\xi}+w)$. Since $(\pi_B\,\hat{\xi}+w)/\Lambda\in(I-f_{t_*})\,\Gamma^b(\Lambda)\cap\Gamma^0(\Lambda)$ it follows that $(\pi_{A_0}\hat{\xi}-w)/\Lambda\in(A/\Lambda)\cap(I-f_{t_*})\,\Gamma^b(\Lambda)$. Let $C_1>0$ be such that every $u\in(A/\Lambda)\cap(I-f_{t_*})\,\Gamma^b(\Lambda)$ has a preimage v such that $\|v\|_0\leqq C_1\,\|u\|_0$. Then take $v\in\Gamma^b(\Lambda)$ satisfying:

$$\begin{aligned}&(I-f_{t_*})\,v=(\pi_{A_0}\hat{\xi}-w)/\Lambda,\\&\qquad\|v\|_0\leqq C_1\,\|\pi_{A_0}\hat{\xi}-w\|_0.\end{aligned}$$

Take $\eta\in\Gamma^b(\Lambda)$ satisfying:

$$\begin{aligned}&(I-f_{t_*})\,\eta=(\pi_B\,\hat{\xi}+w)/\Lambda,\\&\qquad\|\eta\|_0\leqq K\,\|\pi_B\,\xi+w\|_0\end{aligned}$$

where K is given by Proposition 2.1. Then:

$$(I-f_*)(v+\eta)=\xi$$

and

$$\begin{aligned}\|v+\eta\|_0&\leqq C_1\,\|\pi_{A_0}\hat{\xi}-w\|_0+K\,\|\pi_B\,\hat{\xi}+w\|_0\leqq 2C_1\,\|\pi_{A_0}\|\cdot\|\xi\|_0\\&\quad+C_1\,\|w\|_0+2K\,\|\pi_B\|\,\|\xi\|_0+K\,\|w\|_0\leqq(2C_1\,\|\pi_{A_0}\|\\&\quad+k_3\,C_1\,\|\pi_B\|+2K\,k_3\,\|\pi_B\|+2K\,\|\pi_B\|)\,\|\xi\|_0.\end{aligned}$$

Defining C as the factor between brackets the theorem is proved.

Now suppose that $f\colon V\to\mathrm{Diff}^r(M)$ is just a C^k map. If $f_t=I$ we can take $g\in\mathcal{U}$ such that the set of fixed points of g_t is finite for every t. Since $g_{\bar{t}}h=hf_t/\Lambda$ it follows that $h(\Lambda)$ is contained in the set of fixed points of $g_{\bar{t}}$. Hence $h(\Lambda)$ and Λ

are finite. Therefore $\Gamma^b(\Lambda)$ is finite dimensional and the Proposition holds. If $f_t \neq I$ we claim that there exist a neighborhood U of t in V that is a compact manifold and $\xi \in F^{k,r}(U, M)$ such that the map $\tilde{f} \in F^{k,r}(U, M)$ defined by $\tilde{f}_t = \xi_t f_t \xi_t^{-1}$ is an immersion. If this claim is true is easy to see that $\tilde{f}, t$ and $\xi_t(\Lambda)$ satisfy the hypothesis of Proposition 3.1. If $\xi_{t_*}: \Gamma^b(\xi_t(\Lambda)) \to \Gamma^b(\Lambda)$ is defined as $\xi_{t_*}(\eta) = T\xi_t \circ \eta \circ \xi_t^{-1}$ it follows that $(I - f_{t_*})\, \xi_{t_*} = \xi_{t_*}(I - \tilde{f}_{t_*})$. Since ξ_{t_*} is an isomorphism and $(I - \tilde{f}_{t_*})\, \Gamma^b(\xi_t(\Lambda))$ contains a closed finite codimensional subspace of $\Gamma^b(\xi_t(\Lambda))$ we conclude that $(I - f_{t_*})\, \Gamma^b(\Lambda)$ contains a closed finite codimensional subspace of $\Gamma^b(\Lambda)$.

It remains to prove the claim. If $f_t \neq I$ take a finite set of different points x_i, $1 \leqq i \leqq m$, such that $f_t(x_i) \neq x_i$ for every $1 \leqq i \leqq m$. Consider the map $\varphi: V \to M \times \overset{m}{\cdots} \times M$ defined by

$$\varphi(u) = (f_u(x_1), \ldots, f_u(x_m)).$$

If m is large (more precisely $m \dim M \geqq 2 \dim V$) we can approximate φ in the C^r topology by a C^k immersion $\tilde{\varphi}$. Take a neighborhood U of x that is a compact manifold and a C^k map $\xi: U \to \mathrm{Diff}^r(M)$ satisfying:

$$(\xi_u f_u(x_1), \ldots, \xi_u f_u(x_m)) = \tilde{\varphi}(u),$$
$$\xi_u(x_i) = x_i \qquad 1 \leqq i \leqq m$$

when $u \in U$. We want to show that the C^k map $\tilde{f}: u \mapsto \xi_u f_u \xi_u^{-1}$ is an immersion of V in $\mathrm{Diff}^r(M)$. For this define $\Phi: \mathrm{Diff}^r(M) \to M \times \overset{m}{\cdots} \times M$ by $\Phi(g) = (g(x_1), \ldots, g(x_n))$. This map is C^∞ and its composition with $\tilde{f}$ satisfies

$$\begin{aligned}\Phi(\tilde{f}_u) &= (f_u(x_1), \ldots, f_u(x_m)) \\ &= (\xi_u f_u \xi_u^{-1}(x_1), \ldots, \xi_u f_u \xi_u^{-1}(x_m)) \\ &= (\xi_u f_u(x_1), \ldots, \xi_u f_u(x_m)) = \varphi(u)\end{aligned}$$

that is an immersion. Hence $\tilde{f}$ is an immersion.

References

1. Bowen, R.: ω-Limit sets for Axiom A diffeomorphisms Journal of Diff. Eq. Vol. **18**, 333–339 (1975)
2. Brunovsky, P.: Generic properties of the rotation number of one parameter diffeomorphisms of the circle, Czech. Math. Journal, Vol. **24**, 74–90 (1974)
3. Dankner, A.: Axiom A dynamical systems cycles and stability (to appear)
4. Franke, J., Selgrade, J.: Hyperbolicity and chain recurrence, Journal Diff. Eq.
5. Franks, J.: Differentiably Ω-stable diffeomorphism Topology, Vol. **11**, 107–113 (1972)
6. Guckenheimer, J.: Absolutely Ω-stable diffeomorphisms Topology **11**, 195–197 (1972)
7. Hirsch, M., Palis, J., Pugh, C., Shub, M.: Neighborhoods of hyperbolic sets, Invent. Math. **9**, 133–163 (1969/70)
8. Newhouse, S., Palis, J.: Bifurcations of Morse-Smale dynamical systems, Proc. of Symp. on Dynamical Systems (M.M. Peixoto, ed.). New York and London: Academic Press 1973
9. Newhouse, S., Palis, J.: Cycles and bifurcation theory, Asterisque, Soc. Math. France, Vol. 31 (1976) 43–140
10. Newhouse, S., Palis, J., Takens, F.: Stable arcs of diffeomorphisms, Bull, Amer. Math. Soc. **82**, 497–499 (1976)

11. Mañé, R.: Characterization of *AS* diffeomorphisms Proc. of the III Latinoamerican School of Mathematics, Lecture Notes. Notes on Mathematics **597**, 389–395. Berlin-Heidelberg-New York: Springer-Verlag
12. Mañé, R.: Quasi-Anosov diffeomorphisms and hyperbolic manifolds, Trans. Amer. Math. Soc. **229**, 351–370 (1977)
13. Palis, J.: A differentiable Invariant of Topological Conjugacies and Moduli of Stability, Societé Math. de France, Asterisque **51**, 335–346 (1977)
14. Robinson, C.: Global structural stability of the saddle node bifurcation, Trans. Amer. Math. Soc. Vol. **236**, 155–173 (1978)
15. Robinson, C.: Structural stability on manifolds with boundary (to appear)
16. Robinson, C., Williams, R.: Finite stability is not generic, Proc. of Symp. on Dynamical Systems (M.M. Peixoto, ed.). New York and London: Academic Press 1973
17. Selgrade, J.: Isolated invariant sets for flows on vector bundles, Trans. Amer. Math. Soc. **203**, 359–390 (1975)
18. Sotomayor, J.: Generic one parameter families of vectorfields, vectorfields, Publ. IHES Vol. 43, 1974
19. Takens, F.: Singularities of vectorfields, IHES Publ. Math., **43**, 47–100 (1973)

Received November 25, 1978/February 8, 1979

Inventiones math. 53, 77–90 (1979)

Inventiones mathematicae

Associated Curves and Plücker Formulas in Grassmannians

Giuseppe Canuto

Istituto di Geometria, Università di Torino, via Carlo Alberto 10, I-Torino, Italy

Introduction

Classically the study of the extrinsic geometry of a projective curve $f\colon C \to \mathbb{P}^n$ i.e. of a compact Riemann surface C with a holomorphic map into $\mathbb{P}^n(\mathbb{C})$, is done by introducing the associated curves of the original curve. The k^{th} associated curve is the map f_k of C into the Grassmannian $G(k+1; n+1)$ of $(k+1)$-dimensional linear subspaces of $\mathbb{C}^{n+1}$ defined by $f_k(p)=$ the osculating $\mathbb{P}^k$ at p. In this paper we prove that for a generic non-special linear system g_d^n the maps f_k, associated to the map f defined by g_d^n are smooth for $k=1, 2, \ldots, n-2$.

Furthermore the singularities of f_{n-1} can be interpreted as intersection numbers on $J(C)$.

In the second part of the paper we consider curves in Grassmannians. After defining associated curves in analogy with the projective case, we generalize the classical Plücker formulas and discuss briefly the local invariants which arise in a natural way for maps $\phi\colon C \to G(\rho; n)$.

Notations. $C^{(k)}$ is the k-fold symmetric product of C and $\mu\colon C^{(k)} \to J(C)$ is the map into the Jacobian defined by extending by linearity the usual period map from C into $J(C)$. $G(\rho; n)$ is the Grassmannian of ρ-dimensional linear subspaces of $\mathbb{C}^n$ which we always consider as embedded in $\mathbb{P}(\bigwedge^{\rho} \mathbb{C}^n)$ by the Plücker embedding. $G(\rho; n)$ is also the space of $(\rho-1)$-dimensional subspaces of $\mathbb{P}^{n-1}$ and if $e_1, \ldots, e_\rho$ are linearly independent vectors in $\mathbb{C}^n$, $e_1 \wedge e_2 \wedge \ldots \wedge e_\rho$ denotes their span in $\mathbb{C}^n$ or its projectivization according to the context.

By the degree of a curve in $G(\rho; n)$ we mean the degree in $\mathbb{P}(\bigwedge^{\rho} \mathbb{C}^n)$ after the Plücker embedding.

I want to thank Phillip Griffiths and Joe Harris for their help and encouragement during the preparation of this paper.

0020-9910/79/0053/0077/$02.80

1. Projective Curves

Let $f\colon C \to \mathbb{P}^n$ be a map of a compact Riemann surface into projective space. In a neighborhood of a point $p \in C$ we can express the map f with a vector valued function $f(z) = [x_0(z), \ldots, x_n(z)]$. The k^{th} associated curve of f is the map

$$f_k\colon C \to G(k+1; n+1) \subset \mathbb{P}\big(\textstyle\bigwedge^{k+1} \mathbb{C}^{n+1}\big)$$

defined by

$$f_k(z) = [f(z) \wedge f'(z) \wedge \ldots \wedge f^{(k)}(z)].$$

This map is well defined because $f_k(z)$ is independent of the choice of the lifting of f to $\mathbb{C}^{n+1}$ and local coordinate z. The geometrical meaning of $f_k(z)$ is that of osculating k-plane i.e. the unique k-plane in $\mathbb{P}^n$ having contact of order at least $k+1$ with $f(C)$ at z.

For any map $f\colon C \to \mathbb{P}^n$, expressed in terms of Euclidean coordinates by $f_1(z), \ldots, f_n(z)$ the ramification index $\beta(p)$ is defined to be the order of vanishing of the Jacobian matrix

$$\left(\frac{\partial f_1}{\partial z}, \ldots, \frac{\partial f_n}{\partial z}\right) \quad \text{at } p.$$

Choosing a suitable system of coordinates in $\mathbb{C}^{n+1}$ we can express the vector valued function $f(z)$ in a neighborhood of p as

$$\begin{aligned} x_0(z) &= 1 + \ldots \\ x_1(z) &= z^{1+\alpha_1} + \ldots \\ &\vdots \\ x_n(z) &= z^{n+\alpha_1+\ldots+\alpha_n} + \ldots . \end{aligned}$$

This is called the normal form of f around p and the ramification index $\beta_k(p)$ of the k^{th}-associated curve at p is

$$\beta_k(p) = \alpha_{k+1}.$$

(For all this material the reader is referred to [GH] Chap. 2.)

The total ramification of f_k is

$$\beta_k = \sum_{p \in C} \beta_k(\mathrm{p}).$$

When we fix our attention at a point p we sometimes write β_k for $\beta_k(p)$: it will be clear from the context if we refer to the total ramification or the ramification at p.

We can now prove:

Theorem. *Let g_d^r be a generic non-special linear system without base points of dimension r and degree d on a compact Riemann surface C of genus g and $f\colon C \to \mathbb{P}^r$ the corresponding map. Then*

(i) *the first* $(r-2)$ *associated maps of* f *are smooth at each point of* C

(ii) *the map* f_{r-1} *has total ramification*

$$\beta_{r-1}=(r+1)(d+rg-r).$$

More precisely there are β_{r-1} *points* p_i $i=1, \ldots, \beta_{r-1}$ *such that*

$$\alpha_r(p_i)=1 \quad \text{and} \quad \alpha_r(p)=0$$

for any other point p *of* C.

Proof. Let us consider first the case of a complete linear system $|D|$ associated to a non-special divisor D: we have then $d=n+g$.

If $f(z)=[x_0(z), \ldots, x_n(z)]$ is the normal form of the map f at the point p, we see that a hypersurface $a_0 x_0+\ldots+a_n x_n=0$ of $\mathbb{P}^n$ contains $f(p)$ when

$$a_0(1+\ldots)+a_1(z^{1+\alpha_1}+\ldots)+\ldots+a_n(z^{n+\alpha_1+\ldots+\alpha_n}+\ldots)$$

is equal to zero for $z=0$ i.e. when $a_0=0$. Such a hyperplane contains $f(p)$ twice when

$$a_1((1+\alpha_1)\, z^{\alpha_1}+\ldots)+\ldots+a_n((n+\alpha_1+\ldots+\alpha_n)\, z^{n-1+\alpha_1+\ldots+\alpha_n}+\ldots)$$

is zero for $z=0$. This is always true if $\alpha_1 \neq 0$ and it happens only when $a_1=0$ if $\alpha_1=0$ or equivalently $\beta_0=\alpha_1=0$ if and only if $h^0(D-2p)=h^0(D)-2$. By the same argument

$$\beta_0=\beta_1=\ldots=\beta_{h-2}=0 \qquad \text{iff } h^0(D-kp)=h^0(D)-k \qquad \text{for } k=1, 2, \ldots, h.$$

Since D is non-special this can be expressed by saying that $D-kp$ is still non-special for $k=1, 2, \ldots, h$.

Let W^r_{d-k} be the subset of the Jacobian $J(C)$ consisting of divisors E of degree $d-k$ such that $h^0(E) \geqq r+1$.

If we fix k and let the point p vary on C we get a curve

$$C_k=\{D-kp\}_{p\in C} \quad \text{in } J(C)$$

which does not contain any special divisor if it has empty intersection with W^{n+1-k}_{d-k}. A generic translate of C_k in $J(C)$ is a curve $C'_k=\{D'-kp\}_{p\in C}$ corresponding to a non-special divisor D' of degree d. The theorem of the generic translate due to Kleiman [KL] inplies that if

$$(*) \quad \dim C_k+\dim W^{n+1-k}_{d-k}<\dim J(C),$$

then for a generic D the curve C_k does not intersect W^{n+1-k}_{d-k}. The dimension of W^{n+1-k}_{d-k} is the same as the dimension of $W^0_{(2g-2)-(d-k)}=W^0_{g-n+k-2}$ because we can associate to each divisor E the complementary divisor $K-E$, hence the inequality $(*)$ is satisfied for $k \leqq n+1$. This conclues the proof of (i) for a complete system.

By the same argument the map f_{n-1} is not smooth at p exactly when C_{n+1} and the θ-divisor

$$\Theta = W^0_{g-1} = \{\text{effective divisors of degree } g-1\}$$

intersect at $D-(n+1)p$.

Now the curve C_{n+1} has the same homology class of $\{-(n+1)\,p\}_{p\in C}$ which is the image of C under the endomorphism Ψ of $J(C)$ defined by $\Psi(q) = -(n+1)q$, when we consider C embedded in $J(C)$ by the period map. The endomorphism Ψ induces on $H_2(J(C);\mathbb{Z})$ the multiplication by $(n+1)^2$ so we conclude that (if the intersection is transverse)

$$C_{n+1}\cdot\Theta = (n+1)^2\cdot g$$

because, as well known, Θ cuts out on C a divisor of degree g. The transversality of the intersection, for a generic D, is again a consequence of Kleiman's theorem.

Let us consider now a point p where the only ramified maps is f_{n-1}. We have

$$\alpha_1(p) = \ldots = \alpha_{n-1}(p) = 0$$
$$\alpha_n(p) \neq 0 \quad \text{i.e. } h^0(D-(n+1)\,p) = 1$$

and moreover

$$\alpha_n(p) \geqq 2 \quad \text{iff } h^0(D-(n+2)\,p) = 1.$$

By the Singularity Theorem of Riemann (see [GH] Chap. 2) $D' = D-(n+1)p$ is a simple point of Θ and if $D' = q_1 + \ldots + q_{g-1}$ the projectivization of the tangent space to Θ at D' is the span of $q_1, \ldots, q_{g-1}$ in $\mathbb{P}^{g-1}$, which we denote by $\overline{q_1, \ldots, q_{g-1}}$, when we consider C embedded in $\mathbb{P}^{g-1}$ by the canonical divisor.

On the other hand it is immediate to check that the projectivization of the tangent space to C_{n+1} at D' is the point p on the canonical curve.

We can prove now that C_{n+1} and Θ intersect transversally at D' iff $\alpha_n(p) = 1$ which proves (ii) for complete systems. In fact, if the intersection is transverse $p \notin \overline{q_1, \ldots, q_{g-1}}$. If $\alpha_n(p) > 1$, $q_1 + q_2 + \ldots + q_{g-1} - p$ would be effective, but this is possible only if $p = q_i$ for some i contradicting the hypothesis. Suppose on the other hand that $p \in \overline{q_1, \ldots, q_{g-1}}$. Two cases are possible a priori

(a) $p = q_i$ for some i, say $p = q_i$ then $h^0(D-(n+2)\,p) = 1$ and $\alpha_n(p) > 1$.

(b) $p \neq q_i$ for any i then $D - np$ is linearly equivalent to $q_1 + \ldots + q_{g-1} + p$ and $h^0(D-np) = 2$ contradicting the fact that $n\cdot p$ imposes n conditions on D. Thus case (b) cannot occur.

Let us consider now a non-special g^r_d. The complete linear system g^n_d associated to g^r_d is the projectivization of $F = H^0(C;[D])$ where D is any divisor of g^r_d, and g^r_d itself is the projectivization of a subspace H of dimension $r+1$ of F. The first $(r-2)$ maps associated to f are smooth at p if $H^0(C;[D-kp])\cap H$ has dimension $r+1-k$ when $k = 1, \ldots, r$.

By the first part of the proof we can assume that the spaces $F_{n+1-k}(p) = H^0(C;[D-kp])$ $k = 1, \ldots, r$ form a flag on $F = F_{n+1}$ for any p. Following [GH] we denote by $\sigma_{a_1, \ldots, a_{r+1}}$ the Schubert cycle

$$\{V \in G(r+1; n+1) \text{ s.t. } V \cap F_{n-r+i-a_i} \geqq i\}$$

where $F_1 \subset F_2 \subset \dots$ is a flag on $\mathbb{C}^{n+1}$. The set of subspaces of dimension $r+1$ of F having intersection of dimension $\geqq r+2-k$ with $F_{n+1-k}(p)$ is then a cycle $\sigma_{1,1,\dots,1,0,0\dots}$ with the first $r+2-k$ indices equal to one and zeros otherwise. The union of these cycles when p varies in C has codimension $r+1-k$ in $G(r+1; n+1)$ hence the first $(r-2)$ associated maps of a generic g_d^r are smooth everywhere.

Finally if $H \cap H^0(C; [D-(r+1)p]$ has dimension $\geqq 2$, then H belongs to $\sigma_{2,2,0,0\dots}$: the union of these cycles has codimension 3 and generically the ramification points of f_{r-1} have $\beta_{r-1}(p) = \alpha_r(p) = 1$. In order to compute the total ramification β_{r-1} we recall first the classical Plücker formulas ([GH] Chap. 2)

$$d_{k-1} - 2d_k + d_{k+1} = 2g - 2 - \beta_k$$

valid for any map $f: C \to \mathbb{P}^n$. If we assume that $\beta_1 = \beta_2 = \dots = \beta_{n-2} = 0$ then we have

$$\begin{aligned}
-2d_{n-1} + d_{n-2} \quad &= 2g - 2 - \beta_{n-1}\\
2d_{n-1} - 4d_{n-2} + 2d_{n-3} &= 2(2g-2)\\
3d_{n-2} - 6d_{n-3} + 3d_{n-4} &= 3(2g-2)\\
&\ \vdots\\
nd_1 - 2nd \quad &= n(2g-2)
\end{aligned}$$

and adding up we get

$$\beta_{n-1} = (n+1)(d + ng - n).$$

In particular for a complete non-special linear system we get $(n+1)^2 \cdot g = \beta_{n-1}$, as we already know, and for a non-complete g_d^r we get the total ramification β_{r-1}. Q.E.D.

In the case of a complete non-special g_d^n such that the first $(n-2)$ associated maps are everywhere smooth we can be more precise about the map f_{n-1}. We have in fact

Proposition. *Let g_d^n be a complete non-special linear system such that $\beta_0 = \dots = \beta_{n-2} = 0$ for any point. Then $\beta_{n-1}(p) = \alpha_n(p)$ is the intersection multiplicity of Θ and C_{n+1} at $D-(n+1)p$.*

Proof. Suppose Θ and C_{n+1} intersect at $D-(n+1)p$. In order to compute the intersection multiplicity we consider the pull-back $v^*\theta$ of the θ-function, via the map

$$\begin{aligned}
&C \xrightarrow{v} J(C)\\
&p \longrightarrow D-(n+1)p.
\end{aligned}$$

If z is a local coordinate around p we claim that

$$\frac{\partial^h (v^*\theta)}{\partial z^h}(p) = 0$$

for $h = 0, 1, \dots, \alpha_n - 1$.

Since the total intersection between C_{n+1} and Θ is $(n+1)^2 \cdot g$, which is equal to $\sum_{p\in C} \alpha_n(p)$, we see that the intersection multiplicity at $D-(n+1)p$ must be $\alpha_n(p)$. Let us put $D_0=D-(n+1)p$ and define k as the maximum among the numbers r such that D_0-rp is effective. Using the normal form of f we check easily that $k=\alpha_n-1$.

If $\omega_i=\Omega_i dz$ is the expression of the holomorphic differentials around p, $\dfrac{\partial^h(v^*\theta)}{\partial z^h}(p)$ is a linear combination of expressions of the form

$$\sum_{i_1\ldots i_m} \frac{\partial^m \theta}{\partial \mu_{i_1}\ldots\partial\mu_{i_m}}(D_0)\,\Omega_{i_1}^{\delta_1-1}(p)\ldots\Omega_{i_m}^{\delta_m-1}(p)$$

which we denote $\partial^m\theta(\Omega^{\delta_1-1},\ldots,\Omega^{\delta_m-1})(p)$, where $m\leqq h$ and $\sum\delta_i=h$.

We want to show that if D_0-kp is effective then all such expressions with $h\leqq k$ are zero. The proof is by induction on k. For $k=0$ the assertion is true for the theorem we proved, and we suppose that it is true for $k-1$. Now we use induction on m. For $m-1$, $\partial\theta(\Omega^{k-1})=0$ because by assumption $kp\in D_0$. If $D_0=k\cdot p+q_{k+1}+\ldots+q_{g-1}$ we consider the point

$$(\underbrace{p,\ldots,p}_{k\text{ times}},q_{k+1},\ldots q_{g-1})\quad\text{in}\quad C^{(g-1)}$$

which is mapped to D_0 by $\mu\colon C^{(g-1)}\to J(C)$. A system of coordinates about this point is $(w_1,\ldots,w_k,z_{k+1},\ldots,z_{g-1})$ where z_j is a coordinate around q_j and w_i is the i^{th} symmetric function in the coordinate around p. The derivative $\dfrac{\partial^m\theta}{\partial w_{\delta_1}\ldots\partial w_{\delta_m}}(0)$, $\sum\delta_i=k$, is zero because $\theta(\mu(C^{g-1}))\equiv 0$, on the other hand it is a linear combination of expression of the form

$$\sum\frac{\partial^s\theta}{\partial\mu_{i_1}\ldots\partial\mu_{i_s}}(D_0)\cdot\left(\frac{\partial^{m_1}\mu_{i_1}}{\partial T(m_1)}\right)(p)\ldots\left(\frac{\partial^{m_s}\mu_{i_s}}{\partial T(m_s)}\right)(p)$$

where $s\leqq m$, $\mu=\mu_{i_1},\ldots,\mu_{i_s}$.

T is a permutation of $(1,\ldots,m)$

$$m=m_1+m_2+\ldots+m_s$$

and

$$\frac{\partial^{m_1}\mu_{i_1}}{\partial T(m_1)}=\frac{\partial^{m_1}\mu_{i_1}}{\partial w_{\delta_{T(1)}}\ldots\partial w_{\delta_{T(m_1)}}}$$

$$\vdots$$

$$\frac{\partial^{m_s}\mu_{i_s}}{\partial T(m_s)}=\frac{\partial^{m_s}\mu_{i_s}}{\partial w_{\delta_{T(m_1+\ldots+m_{s-1}+1)}}\ldots\partial w_{\delta_{T(m)}}}.$$

The derivatives of the functions μ's with respect to the ω's can be expressed with derivatives of the functions Ω's. We have in fact

Lemma. *The derivative* $\dfrac{\partial^m \mu_i}{\partial^{m_1} w_1 \ldots \partial^{m_k} w_k}(0)$ *is equal to* $\Omega_i^{(\gamma-1)}(p)$, *where* $\gamma = \sum i \cdot m_i$, *multiplied by a constant.*

Using this lemma, the proof of which is given below we see that $\dfrac{\partial^m \theta}{\partial w_{\delta_1} \ldots \partial w_{\delta_m}}(0)$ is a linear combination of expressions of the form $\partial\theta(\Omega^{\gamma_1-1}, \ldots, \Omega^{\gamma_s-1})(p)$ with $s \leqq m$ and $\gamma_1 + \ldots + \gamma_s = k$. If $s \leqq m-1$ all these expressions are equal to zero by induction. If $s = m$ the only expression of this type is

$$\partial^m \theta(\Omega^{\delta_1-1}, \ldots, \Omega^{\delta_m-1})(p)$$

hence also this expression must be zero. Q.E.D.

Proof of the Lemma. Suppose $t(z)$ is an analytic function of z. Define $\mu(z_1, \ldots, z_n) = \sum_{i=1}^{n} t(z_i)$ and let $w_1, \ldots, w_n$ be the first n elementary symmetric functions in $z_1, \ldots, z_n$.

We want to compute $\dfrac{\partial^\alpha \mu}{\partial^{\alpha_1} w_1 \ldots \partial^{\alpha_n} w_n}(0)$; $\alpha = \alpha_1 + \ldots + \alpha_n$. If $t(z_i) = \sum_k \dfrac{t^k(0)}{k!} z_i^k$ then $\mu(z_1, \ldots, z_n) = \sum_k \dfrac{t^k(0)}{k!} s_k$ where $s_k = \sum_i z_i^k$. Using Girard's formula ([MS] p. 195)

$$s_k = \sum_{\beta=k} (-1)^{\alpha+\beta} \frac{(\alpha-1)!\,k}{\alpha_1! \ldots \alpha_n!} w_1^{\alpha_1} \ldots w_n^{\alpha_n}$$

where $\beta = \sum_{i=1}^{n} i \cdot \alpha_i$, we get

$$\mu(z_1, \ldots, z_n) = \sum \frac{t^k(0)}{k!} \left(\sum_{\beta=k} (-1)^{\alpha+\beta} \frac{(\alpha-1)!\,k}{\alpha_1! \ldots \alpha_n!} w_1^{\alpha_1} \ldots w_n^{\alpha_n} \right).$$

Equating this expression to the expansion of μ as power series in $w_1, \ldots, w_n$ we have

$$\frac{\partial^\alpha \mu}{\partial^{\alpha_1} w_1 \ldots \partial^{\alpha_n} w_n}(0) = (-1)^{\alpha+\beta} \frac{(\alpha-1)!}{(\beta-1)!} t^{(\beta)}(0).$$

If we apply this formula to the functions

$$t_i(z) = \int_{p_0}^{z} \omega_i$$

$$\mu(z_1 \ldots z_{g-1}) = \mu_i(z_1, \ldots, z_{g-1})$$

the lemma is proved. Q.E.D.

Special Divisors

In the case of special divisors the theorem on the behavior of the associated curves is not valid. In fact, as we see in the following example, we can find curves having only one (special) linear system g_d^r, thus the generic translate of the curves C_k does not correspond any more to a special divisor of degree d and dimension r. On the other hand the associated maps of the map given by g_d^r do not behave according to the previous theorem.

Consider the intersection C of a fixed quadric surface S and a quartic surface in $\mathbb{P}^3$.

The curve C has degree 8, genus 9 and is embedded in $\mathbb{P}^3$ by a linear system g_8^3 which is complete by Clifford's theorem and special. The canonical divisor is $K_C=2H$ where H is the hyperplane section and since C is contained in only one quadric by degree reasons, the map

$$H^0(\mathbb{P}^3; 2H)\to H^0(C; K_C)$$

is surjective.

We want to see that there is only one g_8^3 on C.

Suppose in fact $D=p_1+\ldots+p_8$ is an effective divisor of projective dimension 3 (any g_8^3 is complete by Clifford).

It is easy to see that $2D=K_C$ but we want to prove that $p_1, \ldots, p_8$ lie on a plane i.e. $D=H$.

We assume for simplicity that the points $p_1, \ldots, p_8$ are distinct.

Since $h^0(K_C-D)=4$ and C is contained in one quadric of $\mathbb{P}^3$, the points of D impose 5 independent conditions on the quadrics of $\mathbb{P}^3$. This means that there are 5 points say p_1, p_2, p_3, p_4, p_5 such that any quadric of $\mathbb{P}^3$ containing them contains p_6, p_7, p_8 also. Our claim will be proved if we show that $p_1, \ldots, p_5$ are contained in a plane.

Suppose this is false.

If four of them, say p_1, p_2, p_3, p_4, lie in a plane α but $p_5 \notin \alpha$, taking quadrics of the form $\alpha+\beta$ where β is a variable plane containing p_5 we see that p_6, p_7, p_8 must lie in α. But then the cone with vertex p_5 over a conic containing p_1, p_2, p_3, p_4 but not p_6 is a quadric which violates the previous conclusion about quadrics through $p_1, \ldots, p_5$.

Suppose now that only three, say p_1, p_2, p_3 lie in a plane α, then the line $\ell_{45} = \overline{p_4 p_5}$ intersects α in a point different from $p_1, \ldots, p_5$ and at least two lines, say $\ell_{12}=\overline{p_1 p_2}$ and $\ell_{13}=\overline{p_1 p_3}$, among the lines connecting $p_1 p_2 p_3$ do not contain the point $\alpha\cdot\ell_{45}$. For any line ℓ containing p_3 and disjoint from ℓ_{12}, ℓ_{45} the quadric

$$Q=\{\text{lines in } \mathbb{P}^3 \text{ meeting } \ell, \ell_{12}, \ell_{45}\}$$

contains p_6, p_7, p_8.

Since ℓ varies we see that p_6, p_7, p_8 must lie in $\ell_{12}\cup\ell_{45}$.

Arguing in the same way with lines through p_2 we see that in fact p_6, p_7, p_8 lie on ℓ_{45}. But this is impossible because in this case ℓ_{45} would be a line of the original quadric Q meeting C, which is of type (4, 4) on Q, in five points.

Let us look now at the inflectionary behavior of such a curve C. Fix a point p on Q.

If $s=[s_0:s_1]$ and $t=[t_0:t_1]$ are the lines of Q passing through p, then $Q\simeq \mathbb{P}^1\times\mathbb{P}^1=s\times t$.

The intersection with quartic surfaces correspond then to polynomials of degree 8 homogeneous of degree 4 in $[s_0:s_1]$ and $[t_0:t_1]$.

Varying the quartic surface we can make the intersection curve tangent at the origin to one of the two lines, say t, and impose a contact of order >2.

As a final remark we notice that in the case of special divisors we can still ask for the generic inflectionary behavior if we allow the curve to vary in some family and we illustrate this with the example of the canonical divisor for curves of genus g. Let C be a curve of genus g. For any point p there are integers

$$1=a_1<a_2<\ldots<a_g \quad \text{s.t.}$$
$$h^0(a_i\cdot p)=h^0((a_i-1)\cdot p)$$

for all i. (See [G] Chap. 7 or [GH] Chap. 2.) These numbers are called the Gap values of p and a point p is regular if the gap sequence is 1, 2, 3, ..., g, otherwise a Weierstrass point. The weight of a Weierstrass point is defined to be

$$W(p)=\sum(a_i-1).$$

A Weierstrass point is hyperelliptic if the gap sequence is

$$a_i=2i-1$$

and C is hyperelliptic iff it contains an hyperelliptic Weierstrass point. In this case all the Weierstrass points are hyperelliptic. If the gap sequence at p is

$$1,2,\ldots,g-1,g+1$$

the p is called a normal Weierstrass point. Now suppose C is not hyperelliptic. If Ω: $C\to\mathbb{P}^{g-1}$ is the canonical map, then the gap values at p are

$$a_h=h+\alpha_1+\ldots+\alpha_{h-1}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_{g-1}$ are the exponents of the normal form of Ω at p.

It follows that

$$W(p)=\sum_{k=1}^{g-1}(g-k)\,\alpha_k$$

and p is a Weierstrass point iff one of associated map of the canonical map is not smooth at p.

The total weight W of C is the sum of $W(p)$, p varying in C, and is not difficult to see that

$$W=(g-1)\,g(g+1).$$

In particular the number of Weierstrass points is finite and it is equal to W when they are all normal.

We have the following

Theorem. *The generic Riemann surface of genus $g \geqq 3$ has only normal Weierstrass points.* (For a proof see [GH] Chap. 2.)

For $g=0,1$ there are no Weierstrass points, for $g=2$ the curve is hyperelliptic. In other words the canonical system which is special has generically the behavior described before for non-special linear systems. In particular for $d=2g-2$ and $n=g-1$ the Plücker formulas give

$$\beta_{n-1}=(g-1)(g+1)$$

which is the total weight.

2. Curves in Grassmannians

Let C be a compact Riemann surface and $\phi: C \to G(\rho; n)$ a holomorphic map into the Grassmannian of ρ-dimensional subspaces of $\mathbb{C}^n$. Locally ϕ is given by choosing ρ holomorphic vectors $f_1(z), \dots, f_\rho(z)$ which span $\phi(z)$ and we write $\phi(z)=f_1(z) \wedge \dots \wedge f_\rho(z)$. Following the example of curves in projective spaces we can construct maps $\phi_k: C \to G((k+1)\rho; n)$ associated to ϕ, in the following way. If $\Lambda_0=f_1 \wedge \dots \wedge f_\rho$ represents locally $\phi=\phi_0$, the map ϕ_k is represented locally by

$$(*) \quad \Lambda_k=(f_1 \wedge \dots \wedge f_\rho) \wedge (f_1' \wedge \dots \wedge f_\rho') \wedge \dots \wedge (f_1^{(k)} \wedge \dots \wedge f_\rho^{(k)}).$$

The definition makes sense because if $f_1^1, \dots, f_\rho'$ are vectors gotten by acting on $f_1, \dots, f_\rho$ with $A \in GL(\rho; \mathbb{C})$ the expression $(*)$ becomes $\Lambda_k^1=(\det A)^{k+1} \Lambda_k$ and if $(\Lambda_k)_\beta$, $(\Lambda_k)_\alpha$ are the expressions of Λ_k in two coordinates z_β, z_α we have

$$(\Lambda_k)_\beta=(\Lambda_k)_\alpha \left(\frac{\partial z_\alpha}{\partial z_\beta}\right)^{\rho \frac{k(k+1)}{2}}.$$

We notice that Λ_k could be identically zero for some k, but in the sequel we consider only non-degenerate curves i.e. such that Λ_k is not identically zero for all the possible k's.

Using the associated maps we can get formulas analogous to the classical Plücker formulas for a curve in projective space.

We recall first some facts about line bundles on a Riemann surface C.

Suppose we are given a family of functions $\{f_\alpha\}_{\alpha \in I}$ on an open covering with coordinate patches $(U_\alpha; z_\alpha)$ of C, s.t.

$$(1) \quad f_\alpha=\|z_\alpha\|^{2\nu} f_\alpha^1$$

where f_α^1 is C^∞ and positive

$$(2) \quad f_\beta=f_\alpha \left\|\frac{\partial z_\alpha}{\partial z_\beta}\right\|^{2k}.$$

If $D=\sum_{p\in C} \nu_p\cdot p$ and T' is the holomorphic tangent bundle of C, by a classical lemma of Kodaira ([K], lemma on p. 1271) the first Chern class of $\overset{k}{\bigotimes} T'\otimes[D]$ is

$$c_1(\overset{k}{\bigotimes} T'\otimes[D])=\left[\frac{i}{2\pi}\,\bar\partial\partial\log f_\alpha^1\right]=\left[\frac{i}{2\pi}\,\bar\partial\partial\log f_\alpha\right].$$

Consider now for each coordinate patch $(U_\alpha;\, z_\alpha)$ of C expressions of the form

$$\Omega_\alpha=\left(\prod_{i=1}^{p}\|\Lambda_{k_i}\|^2\right)\left(\prod_{j=1}^{q}\|\Lambda_{l_j}\|^{-2}\right).$$

If $\sum_i(k_i+1)=\sum_j(l_j+1)$ the previous expressions define positive C^∞ functions with zeros at the zeros of Λ_{k_i} and poles and the zeros of Λ_{l_j}. On the other hand they transform according to the rule

$$\Omega_\beta=\Omega_\alpha\left\|\frac{\partial z_\alpha}{\partial z_\beta}\right\|^{\rho(K-L)}$$

where $K=\sum_i k_i(k_i+1)$ and $L=\sum l_j(l_j+1)$.

Hence if D_{k_i} (resp. D_{l_j}) is the set of zeros of Λ_{k_i} (resp. Λ_{l_j}) and $D_k=D_{k_1}+\ldots+D_{k_p}$ (resp. $D_l=D_{l_1}+\ldots+D_{l_q}$) we have

$$\begin{aligned} &d\Big(\overset{\rho\frac{(K-L)}{2}}{\bigotimes} T'\otimes[D_k]\otimes[-D_l]\Big)\\ &\quad=\sum_{i=1}^{p}\frac{i}{2\pi}\int_C\bar\partial\partial\log\|\Lambda_{k_i}\|^2-\sum_{j=1}^{q}\frac{i}{2\pi}\int_C\bar\partial\partial\log\|\Lambda_{l_j}\|^2.\end{aligned}$$

If d_h is the degree of the h^{th}-associated curve and $\operatorname{ord}(\Lambda_h)$ is the number of zeros of Λ_h the previous equality gives:

$$\rho\,\frac{(K-L)}{2}\,(2-2g)+\sum_{i=1}^{p}\operatorname{ord}(\Lambda_{k_i})-\sum_{j=1}^{q}\operatorname{ord}(\Lambda_{l_j})=\sum_{j=1}^{q}d_{l_j}-\sum_{i=1}^{p}d_{k_i}.$$

If we consider maps $\phi\colon C\to\mathbb{P}^m\simeq G(1;m+1)$ into projective space and we take

$$\Omega=\|\Lambda_{k-1}\|^2\,\|\Lambda_{k+1}\|^2\,\|\Lambda_k\|^{-4}$$

we get

$$(2-2g)+\operatorname{ord}(\Lambda_{k-1})+\operatorname{ord}(\Lambda_{k+1})-2\operatorname{ord}(\Lambda_k)=2d_k-d_{k-1}-d_{k+1}.$$

Using the normal form for maps in projective space it is not difficult to see that the order of vanishing of Λ_k at p is $h\alpha_1(p)+(h-1)\,\alpha_2(p)+\ldots+\alpha_h(p)$. (See [GH] Chap. 2.) As a consequence we can write the Plücker formulas for projective curves in classical form:

$$(2-2g)+\alpha_{k+1}=2d_k-d_{k-1}-d_{k+1}\quad\text{ or }\quad(2-2g)+\beta_k=2d_k-d_{k-1}-d_{k+1}.$$

We can give an application of Plücker formulas to maps $\phi: C \to G(n; 2n)$. In this case $\phi(C)$ can be considered as a ruling $S = \bigcup_{p \in C} \phi(p)$ in $\mathbb{P}^{2n-1}$, traced out by the family of $(n-1)$-planes $\phi(p)$, $p \in C$.

Let $f_1(z), \ldots, f_n(z)$ be the holomorphic vectors spanning $\phi(z)$ and suppose that ϕ is nondegenerate i.e. that Λ_1 is not identically zero. In this case Λ_1 is zero at a finite number of points where the $(n-1)$-plane $\phi(z)$ meets the "infinitely near" $(n-1)$-plane $\phi'(z) = f_1'(z) \wedge \ldots \wedge f_n'(z)$. We have

Proposition. *If ϕ is nondegenerate the number of points where Λ_1 vanishes is*

$$N(\Lambda_1) = 2d + n(2g-2)$$

where g is the genus of C and d is the degree of $\phi(C)$ in $\mathbb{P}(\bigwedge^n \mathbb{C}^{2n})$.

Proof. If we take $\Omega = \|\Lambda_1\|^2 \|\Lambda_0\|^{-4}$ then $\rho = n$ and $K - L = 2$ thus by the general Plücker formulas

$$n(2-2g) + N(\Lambda_1) - \operatorname{ord}(\Lambda_0) = 2d + \frac{i}{2\pi} \int_C \bar{\partial}\partial \log \|\Lambda_1\|^2.$$

Since we can always express locally ϕ in such a way that $\Lambda_0(p) \neq 0$ we see that $\operatorname{ord}(\Lambda_0) = 0$.

On the other hand the integral is zero because we are considering the pull-back of the metric on $\mathbb{P}^0$. Q.E.D.

In particular $N(\Lambda_1)$ is always positive unless $g=0$ and $d=n$.

For $g=0$ and $d<n$ the formula gives negative values of $N(\Lambda_1)$, but in these cases $\Lambda_1 \equiv 0$. In fact the degree of S (which is always equal to the degree of $\phi(C)$) is $\leqq n-1$ and the dimension is n, thus S lies in a hyperplane of $\mathbb{P}^{2n-1}$. This means that $f_1, \ldots, f_n$ (and their derivatives) move in a $\mathbb{C}^{2n-1}$ thus $\Lambda_1 \equiv 0$.

We would like to discuss now the local invariants attached to a map $\phi: C \to G(\rho; n)$.

For a projective curve $\phi: C \to \mathbb{P}^m$ the local invariants appearing in the Plücker formulas and associated maps depend only on the exponents of the normal form. In fact if D is the hyperplane section and we define δ_h as the smallest among the numbers δ such that $h^0(D - \delta p) = m + 1 - h$ we have $\delta_h = h + \alpha_1 + \ldots + \alpha_{h-1}$. On the other hand, as already mentioned, the order of vanishing of Λ_h at p is $h\alpha_1 + (h-1)\alpha_2 + \ldots + \alpha_n$.

There is also a third set of invariants we can look at: for any $p \in C$, $\phi_h(p)$ is the unique h-plane of $\mathbb{P}^m$ with maximal order of contact with $\phi(C)$ at p and the contact is of order

$$h + 1 + \alpha_1 + \ldots + \alpha_{h+1}.$$

Thus the three sets of invariants are equivalent and their equivalence is based on the existence of the normal form.

If ϕ is a map of C into $G(\rho; n)$ where $\rho > 1$ we can still consider at a point p the order of vanishing of Λ_h for the various associated maps ϕ_h. The sequence of numbers $h^0(D - k \cdot p)$ that we considered for projective curves has the following generalization. The map ϕ is completely determined by the rank-ρ vector bundle

E on C which is the pull-back via ϕ of the dual of the universal bundle on $G(\rho; n)$ and by the n-dimensional vector space $V \subset H^0(E)$ which is the pull-back of the linear forms on $\mathbb{C}^n$. For any point $p \in C$ we can consider the spaces

$$V_{k \cdot p} = \{\text{sections in } V \text{ vanishing at } p \text{ with order } \geqq k\}$$

and the corresponding sequence $\dim(V_{k \cdot p})$ $k = 1, 2, \ldots$.

Finally since the integral homology of $G(\rho; n)$ is freely generated by the Schubert cycles $\sigma_{a_1 \ldots a_\rho}$ (as defined in the proof of the theorem) we are led to define osculating cycles of type $\sigma_{a_1 \ldots a_\rho}$ and to consider their order of contact at a point p. However for $\rho > 1$ the equivalence among the various invariants does not hold any more as we see in the following example. Let $\phi: C \to G(2; 4)$ be given locally by the holomorphic vectors $f(z)$, $g(z)$. Fix a basis e_1, e_2, e_3, e_4 of $\mathbb{C}^4$ and suppose that the power series of f and g at p are

$$f(z) = e_1 + e_3 z + \left(\sum_{i=1}^{4} \alpha_i e_i\right) z^2 + \ldots$$

$$g(z) = e_2 + \left(\sum_{i=1}^{4} \beta_i e_i\right) z^2 + \left(\sum_{i=1}^{4} \gamma_i e_i\right) z^3 + \ldots.$$

At $z = 0$ we have

$$\begin{aligned}
&\Lambda_0(p) = e_1 \wedge e_2 = \phi(p),\\
&\Lambda_1(p) = 0,\\
&\Lambda_1'(p) = 2\beta_4 \, e_1 \wedge e_2 \wedge e_3 \wedge e_4,\\
&\Lambda_1''(p) = [8(\alpha_3 \beta_4 - \alpha_4 \beta_3) + 6\gamma_4] \, e_1 \wedge e_2 \wedge e_3 \wedge e_4.
\end{aligned}$$

Since $V^* = \mathbb{C}^4 = (e_1, e_2, e_3, e_4)$

$$\begin{aligned}
&\dim V_{1 \cdot p} = 4 - \dim(\operatorname{span}(f(p), g(p))) = 2\\
&\dim V_{2 \cdot p} = 4 - \dim(\operatorname{span}(f(p), g(p), f'(p), g'(p)) = 1\\
&\dim V_{3 \cdot p} = 1 \quad \text{if } \alpha_4 = \beta_4 = 0\\
&\qquad\qquad\;\; 0 \quad \text{otherwise}
\end{aligned}$$

and the order of vanishing of Λ_1 and $\dim(V_{k \cdot p})$ are completely independent.

Moreover if we pose $\alpha_1 = \alpha_2 = \alpha_3 = 0$, $\alpha_4 = 1$; $\beta_1 = \beta_2 = \beta_4 = 0$, $\beta_3 \neq 0$ and $\gamma_1 = \gamma_2 = \gamma_3 = 0$, $\gamma_4 = 1$ we have

$$\Lambda_1 = \Lambda_1' = 0 \text{ at } z = 0 \quad \text{and} \quad \Lambda_1'' = 0 \text{ iff } \beta_3 = \tfrac{3}{4}.$$

The Schubert cycles on $G(2; 4)$ are

$$\begin{aligned}
\sigma_1(\ell) &= \{\text{lines in } \mathbb{P}^3 \text{ meeting the line } \ell\},\\
\sigma_2(p) &= \{\text{lines in } \mathbb{P}^3 \text{ containing the point } p\},\\
\sigma_{11}(\pi) &= \{\text{lines in } \mathbb{P}^3 \text{ contained in the plane } \pi\},\\
\sigma_{21}(p; \pi) &= \{\text{lines in } \mathbb{P}^3 \text{ containing } p \text{ and contained in } \pi\}.
\end{aligned}$$

For each of these we pull-back to C the defining functions and imposing the maximum order of contact we can determine the osculating cycle of that type.

For example the pull-back of the defining function of $\sigma_2(p)$ is $f(z)\wedge g(z)\wedge p$ where we identify $p\in\mathbb{P}^3$ with any vector in $\mathbb{C}^4$ lifting it.

This function is zero for $z=0$ when $p=\alpha e_1+\beta e_2$ for some α, β.

The first derivative $(f'\wedge g+f\wedge g')\wedge(\alpha e_1+\beta e_2)$ is zero for $z=0$ when $e_3\wedge e_2 \wedge\alpha e_1=0$ i.e. when $\alpha=0$. Thus $p=e_2$ and the osculating cycle of type σ_2 is $\sigma_2(e_2)$.

Since the second derivative is different from zero the order of contact is two.

The analogous computations for the other cycles give: the cycle $\sigma_{11}(\pi)$ has maximal order of contact when π is spanned by e_1, e_2, e_3 (contact of order 2).

When $p=e_2$ and $\pi=e_1\wedge e_2\wedge e_3$ we get the osculating cycle of type σ_{21} (contact of order 2).

Finally the osculating cycle of type σ_1 is $\sigma_1(e_1\wedge e_2)$ and the order of contact depends on β_3 i.e. if $\beta_3\neq 1$ the contact is of order 4, if $\beta_3=1$ the contact is of order >4.

Thus if $\beta_3\neq 0; 1$ the osculating cycles and their order of contact are fixed but the order of vanishing of Λ_1 can vary and vice-versa if $\beta_3=0; \frac{3}{4}$ then Λ_1 vanishes of order 2 but the order of contact of the osculating cycle of type σ_1 can vary.

References

[G] Gunning, R.: Lectures on Riemann Surfaces, Princeton: Princeton University Press, 1966

[GH] Griffiths, P., Harris, J.: Principles of algebraic geometry, New York: John Wiley, 1978

[K] Kodaira, K.: On a differential-geometric method in the theory of analytic stacks, Proc. Nat. Acad. Sci. U.S.A. **39**, 1268–1273 (1953)

[KL] Kleiman, S.: The transversality of a general translate, Compos. Math. **28**, 287–297 (1974)

[MS] Milnor, J., Stasheff, J.: Characteristic Classes, Princeton: Princeton University Press, 1974

Received February 5, 1979

Inventiones math. 53, 91 – 116 (1979)

Inventiones mathematicae

Division Values in Local Fields

Robert F. Coleman

Department of Mathematics, Princeton University, Princeton, N.J., U.S.A.

I. Introduction

In his work on cyclotomic fields Kummer observed that various formal operations on power series had number theoretic applications. Perhaps the most striking of these was Kummer's idea of taking logarithmic derivatives of p-adic numbers. After a long period of neglect, various refinements and generalizations of Kummer's idea have recently been used by Iwasawa [I_1] and Wiles [W] to study explicit reciprocity laws, and by Coates and Wiles [C] to study the arithmetic of elliptic curves with complex multiplication. Other applications of Kummer's observation include Iwasawa's explicit descriptions [I_1], [I_2] of the Galois structure of various modules connected with local cyclotomic fields. The aim of the present paper is to begin a deeper and more systematic study of the local analytic theory which underlies these relations between power series and p-adic numbers.

The central result of this paper is a general theorem on the interpolation of division values, proven in Sect. III. In Sect. IV and VI we present two important applications of this theorem which we shall now describe.

Let K be a fixed local field, i.e. a field which is complete with respect to a discrete valuation and whose residue field is finite. Let $\mathcal{O}_K$ be the ring of integers of K. Fix a local parameter π of $\mathcal{O}_K$, and let $\mathfrak{F}$ be a Lubin-Tate formal group, with endomorphism ring $\mathcal{O}_K$, which is associated with π (see [L]). For b in $\mathcal{O}_K$, we write $[b]$ for the endomorphism of $\mathfrak{F}$ given by b, and $\mathfrak{F}_n$ for the kernel of the endomorphism $[\pi^{n+1}]$. Now take H to be a fixed, complete, unramified extension of K, and let φ be the Frobenius element of the Galois group of H over K. We define the tower of fields

$$H_n = H(\mathfrak{F}_n) \qquad (n = 0, 1, \ldots).$$

When $m \geqq n$, we write $N_{m,n}$ for the norm map from H_m to H_n. We now fix a generator $v = (v_n)$ of the Tate module $\varprojlim \mathfrak{F}_n$ as an $\mathcal{O}_K$-module (in other words, v_n is a generator of $\mathfrak{F}_n$ as an $\mathcal{O}_K$-module for each $n \geqq 0$, and $[\pi^{m-n}](v_m) = v_n$ for all

0020-9910/79/0053/0091/$05.20

$m \geqq n$). Let $\mathcal{O}_H$ be the ring of integers of H. As usual, we write $\mathcal{O}_H((T))$ for the ring of formal Laurent series, with finite poles, in an indeterminate T and with coefficients in $\mathcal{O}_H$. The Galois group of H over K operates on $\mathcal{O}_H((T))$ coefficientwise. Finally, if R is a ring, R^* will denote the multiplicative group of invertible elements of R. Our first application of the interpolation theorem is to coherent sequences of norms (finite or infinite). In particular we prove:

Theorem A. *Let $\alpha = (\alpha_n)$ be an element of $\varprojlim H_n^*$, where the inverse limit is taken with respect to the norm maps. Then there exists a unique power series $f_\alpha(T)$ in $\mathcal{O}^H((T))^*$ satisfying*

$$(\varphi^{-n} f_\alpha)(v_n) = \alpha_n \qquad \textit{for all } n \geqq 0.$$

The uniqueness of $f_\alpha(T)$ is obvious from the Weierstrass preparation theorem. The non-trivial part of the proof of Theorem A is establishing the existence of $f_\alpha(T)$. In the special case $H = K = \mathbb{Q}_p$, Theorem A plays a fundamental role in the paper [C], where an ad hoc proof of it is given.

Now assume that H is finite over K. Let λ be the logarithm map of the Lubin-Tate group $\mathfrak{F}$. If we write $\mathfrak{p}_n$ for the maximal ideal of the ring of integers of H_n, it is well known that λ converges on $\mathfrak{p}_n$. Let $T_{n/K}$ denote the trace map from H_n to K. We then define

$$\mathfrak{X}_n = \{a \in H_n \colon T_n(a\lambda(b)) \in \mathcal{O}_K \text{ for all } b \text{ in } \mathfrak{p}_n\}.$$

If $m \geqq n$, it is plain that the trace map $T_{m,n}$ from H_m to H_n maps $\mathfrak{X}_m$ into $\mathfrak{X}_n$. Our second application of the interpolation theorem is to coherent sequences of traces in $\mathfrak{X}_n$. By means of this result we are able to give a characterization of elements of $\mathfrak{X}_n$ which does not depend on λ. We also obtain:

Theorem B. *Let $\alpha = (\alpha_n)$ be an element of $\mathfrak{X}_\infty = \varprojlim \mathfrak{X}_n$, where the inverse limit is taken with respect to the trace maps. Then there exists a unique power series $g_\alpha(T)$ in $\mathcal{O}_H((T))$ satisfying*

$$\pi^{-(n+1)}(\varphi^{-n} g_\alpha)(v_n) = \alpha_n \qquad \textit{for all } n \geqq 0.$$

Conversely, suppose that a power series $g(T)$ in $\mathcal{O}_H((T))$ has the property that

$$T_{m,n}(g(v_m)) = \pi^{m-n}(\varphi^{m-n} g)(v_n)$$

for all $m \geqq n$. Put $\alpha_n = \pi^{-(n+1)}(\varphi^{-n} g)(v_n)$ for each $n \geqq 0$. Then $\alpha = (\alpha_n)$ belongs to $\varprojlim \mathfrak{X}_n$.

In another paper, in which we will discuss Kummer's logarithmic derivative in detail, we will explain the connection between these results and the explicit reciprocity laws of Wiles. (See the remarks at the end of Sect. VI for further comments.) We also point out that, using Theorem 15 of this paper (which is a stronger version of Theorem A) one can prove in a constructive manner, the following fact, which, previously, was only obtainable indirectly via class field theory. Let $d = [H:K]$.

Proposition. *Let* $m \geqq n \geqq 0$, *and let* $\beta \in H_n^*$. *Then* $\beta \in N_{m,n}(H_m^*)$ *if and only if* $N_{n/K}(\beta) \in (\pi^d)(l+\pi^{m+1}\mathcal{O}_K)$. *Moreover,* $N_{n/K}(H_n^*) = (\pi^d)(l+\pi^{n+1}\mathcal{O}_K)$.

This proposition, however, will not be proven here.

I would like to thank John Coates for his encouragement and help throughout the preparation of this paper.

Notation. Throughout, K will be a fixed local field, and Ω will denote a fixed completion of the algebraic closure of K. All extensions of K will be assumed to lie in Ω. Let L be any such extension of K, and $\mathcal{O}_L$ the ring of integers of L. As usual, $L((T))$ and $L[[T]]$ will denote the field of formal Laurent series in T with coefficients in L and poles of finite order at zero, and the subring consisting of formal series without poles, respectively. The rings $\mathcal{O}_L((T))$ and $\mathcal{O}_L[[T]]$ are defined in a similar manner.

We write $|\ |$ for the absolute value on Ω, normalized so that $|a| = q^{-1}$, where a is any parameter of K and q is the order of the residue field of K. Let B denote the open ball of radius one with center the origin in Ω, and let B' be this open ball with the origin deleted. We write

$$L((T))_1$$

for the subset of $L((T))$ consisting of all power series which converge on B'. We shall always assume that $L((T))_1$ is endowed with the "compact-open" topology with respect to B'. In other words, a sequence $\{f_n\}$ in $L((T))_1$ converges to f if and only if, for each closed annulus A around zero in B', and for each $\varepsilon > 0$, there exists a positive integer $N = N(A, \varepsilon)$ such that $|f_n(a) - f(a)| < \varepsilon$ for all a in A and all $n \geqq N$. It is easy to see that, if $\{f_n\}$ converges to f, then the individual coefficients of the power series f_n converge in L to those of f. We recall several well known facts about power series in $L((T))_1$, when viewed as analytic functions. First, if $f = \sum_{i=-\infty}^{\infty} a_i \cdot T^i$ is any element of $L((T))_1$ then for $r = q^{-a}$, where a is a positive rational, we have,

$$\sup_{\substack{|x|=r \\ x \in \Omega}} |f(x)| = \sup_{n=0,1,\ldots} |a_n| \cdot r^n.$$

Second, elements of $L((T))_1$, satisfy the maximum principle, i.e. for any $f \in L((T))_1$ the maximum of f on any set of the form $\{x \in \Omega\colon q^{-a} \leqq |x| \leqq q^{-b}\}$, where $a > b$ are positive rationals, is attained on the set $\{x \in \Omega\colon |x| = q^{-a}$ or $|x| = q^{-b}\}$. Given positive reals r and ε, with $r < 1$, we define $S_L(r, \varepsilon)$ to be the set of all f in $L((T))_1$ such that $|f(x)| < \varepsilon$ for all x in Ω with $|x| = r$. It follows from the maximum principle that the $S_L(r, \varepsilon)$ form a sub-basis of neighborhoods of the origin in $L((T))_1$. Finally, if L is a complete subfield of Ω then $L[[T]]_1$ is a complete topological subgroup of $L((T))_1$, where $L[[T]]_1$ denotes $L((T))_1 \cap L[[T]]_1$ endowed with the restriction topology from $L((T))_1$.

We will have several occasions to take inverse limits of objects indexed by the non-negative integers. Therefore, we adopt the following notation: If $\{A_n, \gamma_{m,n}\}_{0 \leqq n \leqq m}$ is an inverse system, we let $\gamma_{\infty,m}$ denote the natural map from $\varprojlim A_n$ to A_m. If $a \in \varprojlim A_n$ we set $a_m = \gamma_{\infty,m}(a)$.

II. A Galois Structure on $H((T))_1$

Let K be as in the introduction. We shall study Lubin-Tate formal groups, with endomorphism ring $\mathcal{O}_K$. For a detailed discussion of these, see [L]. We simply recall that they arise in the following manner. Fix a local parameter π in $\mathcal{O}_K$. Let $\mathcal{G}_\pi$ denote the set of all formal power series $f(T)$ in $\mathcal{O}_K[[T]]$ satisfying (i) $f(T) \equiv \pi T$ mod degree 2, and (ii) $f(T) \equiv T^q \bmod \pi \mathcal{O}_K$. For each f in $\mathcal{G}_\pi$, there exists a unique formal group law $\mathfrak{F}_f(X, Y)$ in $\mathcal{O}_K[[X, Y]]$ such that $\mathfrak{F}_f(f(X), f(Y)) = f(\mathfrak{F}_f(X, Y))$. The endomorphism ring of $\mathfrak{F}_f$ is naturally isomorphic to $\mathcal{O}_K$. As usual, we write $[a] = [a]_{\mathfrak{F}_f}$ for the formal power series giving the endomorphism of $\mathfrak{F}_f$ corresponding to $a \in \mathcal{O}_K$. In particular, $[\pi] = f$. Finally, if f and g are any two elements of $\mathcal{G}_\pi$, the formal groups $\mathfrak{F}_f$, $\mathfrak{F}_g$ are isomorphic over $\mathcal{O}_K$. From now on, we fix one of these Lubin-Tate formal groups associated with π and K, and denote it simply by $\mathfrak{F}$.

As in the introduction we let $\mathfrak{F}_n$ denote the kernel of the endomorphism $[\pi^{n+1}]$ of $\mathfrak{F}$. Let H denote a fixed complete (not necessarily finite), unramified extension of K. We put

$$H_n = H(\mathfrak{F}_n), \qquad H_\infty = \bigcup_{n \geqq 0} H_n.$$

Let G_n be the Galois group of H_n over H, and G_∞ the Galois group of H_∞ over H. The action of G_∞ on $\mathfrak{F}_\infty = \bigcup_{n \geqq 0} \mathfrak{F}_n$ gives rise to a continuous homomorphism

$$\kappa \colon G_\infty \to U_K, \tag{1}$$

defined by $\sigma(u) = [\kappa(\sigma)](u)$ for all σ in G_∞ and u in $\mathfrak{F}_\infty$; here U_K denotes the group of units of $\mathcal{O}_K$. By Lubin-Tate Theory, κ is, in fact, an isomorphism. Write $R_n = \mathcal{O}_H[G_n]$ for the group ring of G_n with coefficients in $\mathcal{O}_H$. Let $R_\infty = \varprojlim R_n$, the projective limit being taken with respect to the restriction maps. Intuitively, R_∞ may be viewed as the continuous analogue of the ordinary group ring $\mathcal{O}_H[G_\infty]$. There is a natural injection of $\mathcal{O}_H[G_\infty]$ into R_∞, and we identify $\mathcal{O}_H[G_\infty]$ with its image under this injection. Of course, R_∞ is endowed with the natural topology, arising from viewing each R_n as the product of an appropriate number of copies of $\mathcal{O}_H$. Relative to this topology, $\mathcal{O}_H[G_\infty]$ is a dense subset of R_∞.

The general philosophy behind our work is to translate problems about the arithmetic of the field H_∞ into problems about the structure of the power series ring $H((T))_1$. One of the main benefits of this translation is that it enables us to use differentiation to study the arithmetic of H_∞. As mentioned in the Introduction, the origins of this idea go back to Kummer's work on cyclotomic fields. The following result is of fundamental importance for this translation.

Theorem 1. *There exists a unique, continuous, R_∞-module structure on $H((T))_1$ such that, for all f in $H((T))_1$, we have*

$$\sigma(f) = f \circ [\kappa(\sigma)], \qquad a(f) = a \cdot f, \tag{2}$$

for all σ in G_∞ and all a in $\mathcal{O}_K$.

For the proof we will need the following technical result:

Given $f \in H((T))_1$, we define V_f to be the set of all $g \in H((T))_1$ such that $|g(x)| < \sup_{|y|=|x|} |f(y)|$ for all $x \in B'$. Plainly f belongs to V_f. Also let P_m denote the projection from $H((T))_1$ to itself given by associating to a Laurent series, its terms of degree $<m$. This map is obviously continuous and, by the remarks in the introduction, carries V_f into itself.

Lemma 2. *Suppose $f \in H((T))_1$. Then,* (i) *if $f \in S_H(r, \varepsilon)$ for positive reals r and ε, with $r<1$, then $V_f \subseteq S_H(r, \varepsilon)$.* (ii) *$V_f$ is a complete subspace of $H((T))_1$.* (iii) *If g belongs to $T\mathcal{O}_H[[T]]$, and the coefficient of T in g is a unit in $\mathcal{O}_H$, then $f \circ g \in V_f$.* (iv) *If $\{g_i\}$ is a sequence of elements of V_f then $\lim_{i\to\infty} P_m(g_i)=0$ for all integers m if and only if $\lim_{i\to\infty} g_i=0$.*

Proof. Part (i) follows immediately from the definitions.

It is clear that V_f is a closed subspace of $H((T))_1$. Moreover, it is easy to see that V_f is contained in $T^{-N} \cdot H[[T]]_1$, if f is. Since the latter space is complete, (ii) follows. Next, (iii) is an immediate consequence of the fact that g determines an isometry of B'. Finally, for (iv), it is clear from the above that $g_i - P_m(g_i)$ belongs to V_f, so (iv) will follow easily from the assertion: For positive reals r and ε, $r<1$, there exists a positive integer $N=N(f, r, \varepsilon)$ such that $V_f \cap T^N H[[T]]_1 \subseteq S_H(r, \varepsilon)$. This in turn follows from the assertion: For positive reals r and ε, with $r=q^{-a}$, where a is a positive rational, we have, $S_H(r, \varepsilon) \cap T^m H[[T]]_1 \subseteq S_H(s, \varepsilon(s/r)^m)$ for positive real $s<r$. But this is an easy consequence of the maximum principle.

We also need:

Lemma 2a. *Let $\{a_i\}_{i=1}^{\infty}$ be a sequence of distinct elements of B' such that $\prod_{i=1}^{\infty} a_i = 0$ and let $\{g_n\}_{i=1}^{\infty}$ be a sequence of elements of $\mathcal{O}_\Omega[[T]]$. Then if $\lim_{n\to\infty} g_n(a_i)=0$ for all $i \geqq 1$, it follows that $\lim_{n\to\infty} g_n=0$ in $\Omega((T))_1$.*

Proof. Suppose that $\{g_n\}$ does not converge to zero. Then we claim that without loss of generality we may suppose that $|g_n(0)| > \delta$ for some positive real δ, and for all $n \geqq 1$. Indeed, since all the g_n lie in $\mathcal{O}_\Omega[[T]]$, there must exist a smallest k such that $c_{n,k}$ does not converge to zero, where $c_{n,k}$ is the coefficient of T^k in g_n. The conditions of the lemma do not change it we replace g_n by $T^{-k}(g_n - P_k(g_n))$. Our claim is then established by choosing an appropriate subsequence of $\{g_n\}$.

The lemma will follow from the following assertion: Set $A_m = \prod_{i=1}^{m} |a_i| \cdot \prod^{m} |(a_i - a_j)|$ and let $f \in \mathcal{O}_\Omega[[T]]$. Then, if $|f(a_i)| < A_m$, for $1 \leqq i \leqq m$, we have $|f(0)| < \prod_{i=1}^{m} |a_i|$. We prove this assertion by induction on m. If $m=1$ then $|a_1| = A_1 > |f(a_1)| = |f(0) + (f(a_1) - f(0))|$, so as $|f(a_1) - f(0)| < |a_1|$, the assertion follows in this case. Now suppose the assertion true for $m \geqq 1$. Expand f around a_{m+1} so that $f(T) = f(a_{m+1}) + (T - a_{m+1})\, g(T)$ for some $g \in \mathcal{O}_\Omega[[T]]$. Evaluating this expression at a_i and using the hypothesis that $|f(a_i)| < A_{m+1}$, for all $1 \leqq i \leqq m+1$, we find

$$|a_i - a_{m+1}| \cdot |g(a_i)| < A_{m+1}$$

for $1 \leqq i \leqq m$. Since $|a_j| < 1$ and $a_i \neq a_{m+1}$, for $i, j \leqq m+1$, $i \neq m+1$, it follows that

$$g(a_i) < A_{m+1}(|a_i - a_{m+1}|)^{-1} < A_m$$

for all $1 \leqq i \leqq m$. Thus by induction $|g(0)| < \prod_{i=1}^{m} |a_i|$. Therefore $|f(0)| = |f(a_{m+1}) - a_{m+1} \cdot g(0)| < \prod_{i=1}^{m+1} |a_i|$ as asserted.

Let $\mathfrak{F}'_n = \{u \in \mathfrak{F}_n : u \neq 0\}$, for all $0 \leqq n \leqq \infty$. Since $\prod_{u \in \mathfrak{F}'_\infty} u = 0$, Lemma 2a has the following consequence, which will play an important role in the rest of the paper.

Uniqueness Principle. *If f and g are in $\mathcal{O}_H((T))$ and $f(u) = g(u)$ for all $u \in \mathfrak{F}'_\infty$ then $f = g$.*

Proof of Theorem 1. To simplify notation in the proof, we put $M = H((T))_1$. It follows from (iii) of Lemma 2 that $f \circ [\kappa(\sigma)]$ also belongs to M. Thus (2) defines an $\mathcal{O}_H[\mathrm{G}_\infty]$-module structure on M, whence the uniqueness is plain, because $\mathcal{O}_H[G_\infty]$ is dense in R_∞.

We now turn to existence. We conclude from (iii) of Lemma 2 that ωf lies in V_f, for all ω in $\mathcal{O}_H[G_\infty]$. Since V_f is a complete subspace of M, if we can show that

$$\lim_{i \to \infty} \omega_i f = 0, \tag{4}$$

for any sequence $\{\omega_i\}$ of elements of $\mathcal{O}_H[G_\infty]$ such that $\lim_{i \to \infty} \omega_i = 0$, and for any f in M, then we will be able to extend the map $\omega \mapsto \omega f$ by continuity to a continuous map from R_∞ into V_f. This will give the desired action of R_∞ on M. Suppose for the moment that we have already established (4). To show that the map $R_\infty \times M \to M$, given by $(\omega, f) \mapsto \omega f$, is continuous, we need only verify that

$$\lim_{i \to \infty} \omega_i f_i = 0, \tag{5}$$

for any sequence (ω_i, f_i) in $R_\infty \times M$ such that $\lim_{i \to \infty} (\omega_i, f_i) = (\omega, 0)$ or $(0, f)$, for some $\omega \in R_\infty$ or $f \in M$. In the first case, $\lim_{i \to \infty} f_i = 0$, and since $\omega_i f_i \in V_{f_i}$ for all i, (5) follows from (i) of Lemma 2. In the second case, what we have just shown implies that $\lim_{i \to \infty} \omega_i (f_i - f) = 0$, and so (5) follows from the continuity of the map $\omega \mapsto \omega f$.

Thus, to complete the proof, we must establish (4). Suppose first that $f \in \mathcal{O}_H((T))$. Now R_∞ acts continuously on the additive group of H_∞ in the natural way. Thus $\lim_{i \to \infty} \omega_i(b) = 0$ for all $b \in H_\infty$. Taking $b = f(u)$, where u is any non-zero element of $\mathfrak{F}_\infty$, we conclude that $\lim_{i \to \infty} \omega_i(f(u)) = 0$. But $\omega_i(f(u)) = (\omega_i f)(u)$, because $\sigma(u)=[\kappa(\sigma)](u)$ for all σ in G_∞. From this and Lemma 2a, (4) follows in this case. Clearly this also implies that (4) holds for all f in $c.\ \mathcal{O}_H((T))$, where c is any element of H.

Now, for each integer $m \geqq 0$, there exists a $c_m \neq 0$ in H such that $P_m(f)$ is in $c_m \cdot \mathcal{O}_H((T))$. Thus, by what we have already shown $\lim_{i\to\infty} \omega_i(P_m(f))=0$. But, it is easy to see that for any ω in $\mathcal{O}_H[G_\infty]$, we have $P_m(\omega f)=P_m(\omega P_m(f))$. Therefore, the above equation implies that $\operatorname{Lim}_{i\to\infty} P_m(\omega_i f)=0$. This together with Lemma 2 (iv) establishes (4), and the proof of the theorem is complete.

From now on, we shall always consider $H((T))_1$ as being endowed with the R_∞-module structure given by Theorem 1. Plainly, $H[[T]]_1$, $\mathcal{O}_H((T))$ and $\mathcal{O}_H[[T]]$ are then R_∞-submodules of $H((T))_1$.

III. The Trace Operator

We use the same notation as in the previous sections. In particular, H denotes an arbitrary complete, unramified extension of K, and $H_n = H(\mathfrak{F}_n)$, $(n=0, 1 \ldots)$. For brevity, we write $\mathcal{O}_n$ for the ring of integers of H_n, and $\mathfrak{p}_n$ for the maximal ideal of $\mathcal{O}_n$.

We denote the sum of two elements X and Y under the formal group law of $\mathfrak{F}$ by $X[+]Y$. For $f \in H((T))_1$, n a positive integer, $u \in \mathfrak{F}_\infty$ we define the elements f_{π^n} and ${}_u f$ of $H_0((T))$ by

$$f_{\pi^n} = f \circ [\pi^n], \qquad {}_u f(T) = f(T[+]u).$$

(Note that these elements are not necessarily in $H_0((T))_1$.)

As in the introduction, let $v=(v_n)$ be a generator for the Tate-module of $\mathfrak{F}$ as an $\mathcal{O}_K$-module.

Lemma 3. *If $f \in \mathcal{O}_H[[T]]$ and ${}_u f = f$ for all $u \in \mathfrak{F}_0$ then there exists a unique $g \in \mathcal{O}_H[[T]]$ such that $g_\pi = f$.*

Proof. Uniqueness is obvious since $[\pi]$ has a power series inverse in $H[[T]]$.

Now, suppose we have constructed a_i in $\mathcal{O}_H$ for $0 \leqq i \leqq n-1$ so that

$$[\pi]^n \cdot f_n = f - \sum_{i=0}^{n-1} a_i [\pi]^i \tag{1}$$

for some $f_n \in \mathcal{O}_H[[T]]$ (this is trivial for $n=0$). Then as ${}_u f = f$ and ${}_u[\pi]=[\pi]$, for all $u \in \mathfrak{F}_0$, it follows that ${}_u f_n = f_n$. But then $(f_n - f_n(0))(u)=0$ for all u in $\mathfrak{F}_0$, so by Weierstrass preparation there exists an $f_{n+1} \in \mathcal{O}_H[[T]]$ such that $f_n = f_n(0) + [\pi] \cdot f_{n+1}$. Therefore we may set $a_n = f_n(0)$ and obtain (1) for $n+1$. In this manner we may construct a sequence $\{a_i\}$ of elements of $\mathcal{O}_H$ such that

$$f - \sum_{i=0}^{\infty} a_i [\pi]^i \in \bigcap_{n \geqq 0} [\pi]^n \mathcal{O}_H[[T]] = (0).$$

Setting $g = \sum_{i=0}^{\infty} a_i T^i$ we complete the proof of the lemma.

Theorem 4. *There exists a unique map* $\mathcal{S}: H((T))_1 \to H((T))_1$ *such that*

$$\mathcal{S}(f)_\pi = \sum_{u \in \mathfrak{F}_0} {}_u f. \tag{2}$$

Moreover, $\mathcal{S}$ *is continuous.*

Proof. As in Lemma 3 the uniqueness is plain.

For $f \in H((T))_1$ let $\mathcal{T}(f) = \sum_{u \in \mathfrak{F}_0} {}_u f$; $\mathcal{T}(f)$ is then in $H((T))$.

To prove the existence of $\mathcal{S}$, first suppose $f \in \mathcal{O}_H[[T]]$. It follows that $\mathcal{T}(f) \in \mathcal{O}_H[[T]]$ and that ${}_u\mathcal{T}(f) = \mathcal{T}(f)$ for all $u \in \mathfrak{F}_0$. Hence, by Lemma 3, $\mathcal{T}(f) = g_\pi$ for a unique $g \in \mathcal{O}_H[[T]]$. We set $\mathcal{S}(f) = g$ in this case. This defines $\mathcal{S}$ on $\mathcal{O}_H[[T]]$. It is clear that $\mathcal{S}$ is $\mathcal{O}_H$-linear, has its image in $\mathcal{O}_H[[T]]$ and satisfies (2). Therefore, we may extend $\mathcal{S}$ H-linearly to $H \otimes \mathcal{O}^H[[T]]$. Its image is then contained in $H[[T]]_1 = H((T))_1 \cap H[[T]]$ and (2) holds. Since $H \otimes \mathcal{O}_H[[T]]$ is dense in $H[[T]]_1$, if we verify that $\mathcal{S}$ is continuous on this set, then we may extend it by continuity to all of $H[[T]]_1$. (We use here that $H[[T]]_1$ is a complete topological group.) It follows immediately that $\mathcal{S}$ will be continuous, and since $\mathcal{T}$ is evidently continuous, that (2) will hold.

Since $\mathcal{S}$ is H-linear on $H \otimes \mathcal{O}_H[[T]]$ we need only verify continuity at the origin. Let r, ε be positive real numbers with $q^{-(q-1)^{-1}} < r < 1$. It suffices to show: If $f \in S_H(r, \varepsilon) \cap H \otimes \mathcal{O}_H[[T]]$ then $\mathcal{S}(f) \in S_H(r^q, \varepsilon)$. It is clear that ${}_u f$ is in $S_{H^0}(r, \varepsilon)$ if $u \in \mathfrak{F}_0$, since then $|u| < q^{-(q-1)^{-1}}$. Thus $\mathcal{T}(f)$ belongs to $S_H(r, \varepsilon)$. Also, if B_t denotes the ball of radius t around the origin in Ω, then it is easy to see that $[\pi](B_r) = B_{r^q}$. Hence as $\mathcal{S}(f) \in H[[T]]_1$ and satisfies (2) we have: $\mathcal{S}(f)(B_{r^q}) = \mathcal{T}(f)(B_r)$; and the above assertion follows immediately. Therefore, we may extend $\mathcal{S}$ to all of $H[[T]]_1$.

It remains to define $\mathcal{S}$ for f arbitrary in $H((T))_1$. However, if $f \in H((T))_1$, there exists a suitably large positive integer N such that $[\pi]^N \cdot f \in H[[T]]_1$. We then set

$$\mathcal{S}(f) = T^{-N} \mathcal{S}([\pi]^N \cdot f).$$

It is clear that $\mathcal{S}$ will now satisfy the requirements of the theorem.

Recall (cf. Theorem 1) that $H((T))_1$ has a canonical structure as an R_∞-module. If $m \geq n$, $T_{m,n}$ denotes the trace map from H_m to H_n; also recall, $\mathfrak{F}'_n = \mathfrak{F}_n - \{0\}$, for $0 \leq n \leq \infty$.

Corollary 5. *The map* $\mathcal{S}$ *is a continuous* R_∞*-endomorphism of* $H((T))_1$, *which leaves invariant the submodules* $H[[T]]_1$, $\mathcal{O}_H((T))$ *and* $\mathcal{O}_H[[T]]$. *Moreover, for all* f *in* $H((T))_1$, *we have* (i) $\mathcal{S}(f)(v_n) = T_{n+1,n}(f(v_{n+1}))$ *for all* $n \geq 0$; (ii) $\sum_{u \in \mathfrak{F}'_{n-1}} f(u) = (\mathcal{S}^n(f)_{\pi^n} - f)(0)$, $n < \infty$.

Proof. It is clear from the proof of Theorem 4 that $\mathcal{S}$ leaves invariant $\mathcal{O}_H((T))$, $\mathcal{O}_H[[T]]$, and $H[[T]]_1$.

Now

$$\sigma \mathcal{S}(f)_\pi = \sigma \sum_{v \in \mathfrak{F}_0} {}_v f = \sum_{v \in \mathfrak{F}_0} {}_{\sigma^{-1}(v)} \sigma f = \sum_{v \in \mathfrak{F}_0} {}_v \sigma f = \mathcal{S}(\sigma f)_\pi.$$

Thus $\sigma\mathcal{S}(f)=\mathcal{S}(\sigma f)$. Also, $\mathcal{S}$ is obviously H-linear, *a fortiori* $\mathcal{O}_H$-linear, and so it follows from the continuity of $\mathcal{S}$ that $\mathcal{S}$ is an R_∞-homomorphism.

(i) follows immediately from (2) since the conjugates of $f(v_n)$ over H_{n-1} are precisely the elements $f(v_n[+]u)$ of H_n, for $u\in\mathfrak{F}_0$.

By iterating (2) we obtain

$$\mathcal{S}^n(f)_{\pi^n}=\sum_{u\in\mathfrak{F}_{n-1}} {}_u f.$$

If we subtract f from both sides of this expression we cancel the pole and hence we may evaluate at zero to obtain (ii).

The trace operator also enjoys the following congruence property:

Lemma 6. *If* $f\in\mathcal{O}_H((T))$ *then*

$$\mathcal{S}^n(f)\equiv 0 \bmod \pi^n\mathcal{O}_H((T)).$$

Proof. The lemma follows easily from the case $n=1$ using induction and the $\mathcal{O}_H$-linearity of $\mathcal{S}$.

First suppose $f\in\mathcal{O}_H[[T]]$. Then since $f_\pi\equiv f(T^q) \bmod \pi$ and ${}_uf\equiv f \bmod \mathfrak{p}_0$ it follows from (2) that

$$\mathcal{S}(f)(T^q)\equiv q\cdot f\equiv 0 \bmod \mathfrak{p}_0$$

and hence mod π since the left hand side is in $\mathcal{O}_H[[T]]$. This implies the lemma for $f\in\mathcal{O}_H[[T]]$. Now suppose f arbitrary in $\mathcal{O}_H((T))$. There exists a positive integer N such that $[\pi]^N\cdot f\in\mathcal{O}_H[[T]]$. It is easily checked that

$$\mathcal{S}([\pi]^N\cdot f)=T^N\mathcal{S}(f).$$

Hence, the lemma follows immediately from the special case, $f\in\mathcal{O}_H[[T]]$; discussed above.

We now begin the proof of the general interpolation theorem mentioned in the introduction.

Let $\mathcal{H}$ denote the set of all Galois equivariant maps from $\mathfrak{F}'_\infty$ into H_∞ (i.e. w.r.t. G_∞). Then, $\mathcal{H}$ is naturally an R_∞-module where the action is given by $(\omega h)(u)=\omega(h(u))$ for $\omega\in R_\infty$, $h\in\mathcal{H}$ and $u\in\mathfrak{F}'_\infty$. Moreover, restriction from B' to $\mathfrak{F}'_\infty$ defines a continuous R_∞-homomorphism from $H((T))_1$ into $\mathcal{H}$. For a given R_∞-submodule A of $H((T))_1$, two natural questions arise; What is the kernel and what is the image of this homomorphism restricted to A?

If $A=H((T))_1$, the answers are that the kernel is $\lambda\cdot H((T))_1$, where λ is the logarithm of $\mathfrak{F}$ (see Sect. V), and the image is all of $\mathcal{H}$. It is more interesting to consider submodules of $H((T))_1$ defined by some "integrality" condition. Here we shall be mainly concerned with $\mathcal{O}_H((T))$.

Our uniqueness principle answers our first question for $A=\mathcal{O}_H((T))$; that is, the kernel is zero. To answer our second question, we introduce the following integral.

For $h \in \mathscr{H}$, we define

$$\int_{\mathfrak{F}} h = -\operatorname*{Lim}_{n \to \infty} \sum_{u \in \mathfrak{F}'_n} h(u)$$

whenever the limit exists. Let $L(\mathfrak{F})$ denote the subset of $\mathscr{H}$ on which this integral is defined. $L(\mathfrak{F})$ is clearly an R_∞-submodule of $\mathscr{H}$ and $\int_{\mathfrak{F}}$ defines a linear functional from $L(\mathfrak{F})$ into H. Let T_n denote the trace from H_n to H. It follows from the Galois equivariance that if $h \in L(\mathfrak{F})$,

$$\int_{\mathfrak{F}} h = -\sum_{n=0}^{\infty} T_n(h(v_n)) \tag{3}$$

(recall that v_n is a generator for $\mathfrak{F}_n$). The significance of this integral lies in the following mean-value property for elements of $\mathcal{O}_H[[T]]$.

Proposition 7. *If $f \in \mathcal{O}_H[[T]]$ then $f \in L(\mathfrak{F})$ and*

$$f(0) = \int_{\mathfrak{F}} f.$$

Proof. By Lemma 5 (ii) $-\sum_{u \in \mathfrak{F}'_n} f(u) = f(0) - \mathscr{S}^{n+1}(f)(0)$, so

$$\begin{aligned}\int_{\mathfrak{F}} f &= f(0) - \lim_{n \to \infty} \mathscr{S}^{n+1}(f)(0)\\ &= f(0),\end{aligned}$$

using Lemma 6.

For $g \in H((T))_1$ and $h \in \mathscr{H}$ we understand by $g \cdot h$, the element of $\mathscr{H}$ defined as the pointwise product of g and h as functions on $\mathfrak{F}'_\infty$. The previous proposition motivates us to define $\mathscr{H}_n$, $0 \leqq n \leqq \infty$ to be the set of all elements h of $\mathscr{H}$ such that $g \cdot h \in L(\mathfrak{F})$ and

$$\int_{\mathfrak{F}} g \cdot h \equiv 0 \bmod \pi^{n+1} \mathcal{O}_H$$

for all $g \in T\mathcal{O}_H[[T]]$, where we take $\pi^{\infty+1} = 0$.

We now state the main result of this section.

Theorem 8. *If $h \in \mathscr{H}$ and $k \in \mathbb{Z}$, then $T^k \cdot h \in \mathscr{H}_n$, $0 \leqq n \leqq \infty$, if and only if there exists an $f \in T^{-k}\mathcal{O}_H[[T]]$ such that*

$$f(u) = h(u) \quad \textit{for all } u \in \mathfrak{F}'_n.$$

We shall require a couple of lemmas.

Lemma 9. *If $\alpha_i \in \pi^{n-i} \mathfrak{p}_0 \mathcal{O}_n$ for $0 \leqq i \leqq n < \infty$, then there exists an $f \in \mathcal{O}_H[[T]]$ such that $f(v_i) = \alpha_i$.*

Proof. This follows simply from the observation that if

$$g_{n,k} = \frac{[\pi^{n+1}] \cdot [\pi^k]}{[\pi^{k+1}]} \quad \text{for } 0 \leqq k \leqq n,$$

then $g_{n,k} \in \mathcal{O}_H[[T]]$ and

$$g_{n,k}(v_i) = \begin{cases} 0 & 0 \leqq i \leqq n \quad i \neq k \\ \pi^{n-k} \cdot v_0 & i = k. \end{cases}$$

Lemma 10. *If $f \in \mathcal{O}_H[[T]]$ and $T^{-1} f(T) \in \mathcal{H}_n$, $n < \infty$ then there exists a $g \in \mathcal{O}_H[[T]]$ such that $g(u) = u^{-1} f(u)$, for $u \in \mathfrak{F}'_n$.*

Proof. Since $T^{-1} \cdot f(T) \in \mathcal{H}_n$,

$$\int_{\mathfrak{F}} f \equiv 0 \bmod \pi^{n+1} \mathcal{O}_H.$$

But, $\int_{\mathfrak{F}} f = f(0)$ by Proposition 7 and so the above congruence implies $f(0) \in \pi^{n+1} \mathcal{O}_H$. Let $f(0) = \pi^{n+1} b$, $b \in \mathcal{O}_H$ and let

$$g(T) = T^{-1} f(T) - b T^{-2} [\pi^{n+1}](T).$$

By construction, g satisfies the conditions of our lemma.

Proof of Theorem 8. Clearly it suffices to consider only the case $k=0$. First suppose $n < \infty$. Let $h \in \mathcal{H}_n$. On the one hand we observe that $T^r \cdot h \in \mathcal{H}_n$ for all $r \geqq 0$; so if there exists an $r \gg 0$ and an $f' \in \mathcal{O}_H[[T]]$, such that $f'(u) = u^r h(u)$ for $u \in \mathfrak{F}'_n$, then applying Lemma 10 iteratively we deduce that there exists an $f \in \mathcal{O}_H[[T]]$ satisfying the requirements of the Theorem. On the other hand, for r sufficiently large,

$$u_i^r h(u_i) \in \pi^{n-i} \mathfrak{p}_0 \mathcal{O}_i \quad \text{for } 0 \leqq i \leqq n.$$

Hence, by Lemma 9, there exists an f' such that $f'(v_i) = v_i^r \cdot h(v_i)$ for $0 \leqq i \leqq n$, and by Galois equivariance $f'(u) = u^r \cdot h(u)$ for $u \in \mathfrak{F}'_n$. This together with the above observation completes the proof for $n < \infty$.

If $n = \infty$, then by what we have already proven, for each $m \geqq 0$, there exists an $f_m \in \mathcal{O}_H[[T]]$ such that $f_m(u) = h(u)$ for $u \in \mathfrak{F}'_\infty$. By Lemma 2a, $\{f_m\}$ is a Cauchy sequence. Let f be its limit in $\mathcal{O}_H[[T]]$ ($\mathcal{O}_H[[T]]$ is a complete subspace of $H((T))_1$). Then we must have $f(u) = h(u)$ for all $u \in \mathfrak{F}'_\infty$. This concludes the proof of the "if" part of the Theorem. The converse follows immediately from Proposition 7, and the fact that $T_m(\mathcal{O}_m) \subseteq \pi^m \mathcal{O}_H$ (which follows from Lemma 6 and Corollary 5 (i)).

IV. The Norm Operator

To simplify notation we set $\mathcal{O}_H[[T]] = I$. Let $\mathcal{M}$ denote the group of invertible elements in $\mathcal{O}_H((T))$ and $\mathcal{M}^o$ the group of principal units in I, i.e., the set of all $f \in I$ such that $f(0) \equiv 1 \bmod \pi \mathcal{O}_H$. Let p be the characteristic of the residue field of K. Then $I^* = V \times \mathcal{M}^o$ where V is the group of roots of unity of order prime to p in H.

Let $\mathbb{Z}_p$ denote the p-adic integers and let $\mathcal{T}_\infty = \varprojlim \mathbb{Z}_p[G_n]$, where the inverse limit is taken with respect to the canonical restriction maps. $\mathbb{Z}_p$ acts in a natural way on $\mathcal{M}^o$ and we may give $\mathcal{M}^o$ the structure of a continuous $\mathcal{T}_\infty$-module such that

$$(f)^a = f^a \quad \text{and} \quad (f)^\sigma = f \circ [K(\sigma)]$$

for $a \in \mathbb{Z}_p$ and $\sigma \in G_\infty$. The proof of this fact runs along similar lines to the proof of Theorem 1, only in this case it is much simpler as $\mathcal{M}^o$ is complete.

For each $n \geqq 0$, let $\varepsilon_n\colon H((T))_1 \to H_n$ be the evaluation map; $f \mapsto f(v_n)$. It is clear that ε_n induces an R_∞-homomorphism. Also, if we restrict ε_n to $\mathcal{M}^o$ we obtain a $\mathcal{T}_\infty$-homomorphism from $\mathcal{M}^o$ into $1 + \mathfrak{p}_n \subseteq H_n^*$, where $\mathcal{T}_\infty$ acts on the multiplicative group $1 + \mathfrak{p}_n$ in the natural way.

Theorem 11. *There exists a unique map $\mathcal{N}\colon \mathcal{O}_H((T)) \to \mathcal{O}_H((T))$ which satisfies:*

$$\mathcal{N}(f)_\pi = \prod_{u \in \mathfrak{F}_0} {}_u f. \tag{1}$$

Moreover, $\mathcal{N}$ is continuous.

Proof. Uniqueness is plain.

For $f \in I$, let $\mathcal{Q}(f) = \prod_{u \in \mathfrak{F}_0} {}_u f$. It is easy to see that $\mathcal{Q}(f) \in I$ and that ${}_u\mathcal{Q}(f) = \mathcal{Q}(f)$ for all $u \in \mathfrak{F}$. Thus, by Lemma 3, there exists a unique $g \in I$ such that $g_\pi = \mathcal{Q}(f)$. In this case, set $\mathcal{N}(f) = g$. For f arbitrary in $\mathcal{O}_H((T))$ there exists a suitably large positive integer N such that $[\pi]^N \cdot f \in I$, we then set $\mathcal{N}(f) = T^{-q^N} \mathcal{N}([\pi]^N \cdot f)$. It follows easily now that $\mathcal{N}$ satisfies the requirements of the Theorem.

Continuity follows from the continuity of $\mathcal{Q}$ (cf. the proof of Theorem 4).

Let $N_{m,n}$ denote the norm from H_m to H_n, and N_n the norm from H_n to H. Let $v_T(f)$ denote the order of the zero of f at zero, for $f \in H((T))$.

Corollary 12. *The map $\mathcal{N}$ is multiplicative from the monoid $\mathcal{O}_H((T))$ into itself, and leaves invariant $\mathcal{M}$ and $\mathcal{M}^o$. Moreover, if $f \in \mathcal{O}_H((T))$ then;* (i) $v_T(\mathcal{N}(f)) = v_T(f)$. (ii) $\varepsilon_n \mathcal{N}(f) = N_{n+1,n}(\varepsilon_{n+1}(f))$. (iii) *$\mathcal{N}$ restricts to a continuous $\mathcal{T}_\infty$-endomorphism of $\mathcal{M}^o$.*

Proof. That $\mathcal{N}$ is multiplicative follows immediately from the fact that $\mathcal{Q}$ is. It follows that $\mathcal{N}$ leaves invariant $\mathcal{M} = \mathcal{O}_H((T))^x$. From the proof of Theorem 11, we see that $\mathcal{N}$ leaves invariant I, hence also $\mathcal{M} \cap I^x = V \times \mathcal{M}^o$. Since $\mathcal{M}^o$ is a pro-p group and V is a torsion group without p-torsion, it follows that $\mathcal{N}$ leaves invariant $\mathcal{M}^o$. Now, (i) and (ii) follow immediately from (1). Finally, (1) also implies that $\mathcal{N}(h^\sigma) = \mathcal{N}(h)^\sigma$ for $h \in \mathcal{M}^o$; hence (iii) follows from the continuity of $\mathcal{N}$.

Remark. If the characteristic of K is zero and $\mathcal{T}_\infty$ is identified with a subring of R_∞ in the natural way, then the usual logarithm series induces a $\mathcal{T}_\infty$-homomorphism of $\mathcal{M}^0$ into $H[[T]]_1$ such that $\operatorname{Log} \mathcal{N}(h) = \mathcal{S} \operatorname{Log}(h)$.

Let φ be the Frobenius automorphism of H_∞ over K_∞. For $f \in H((T))$ we define φf coefficientwise so that φ commutes with G_∞, $\mathcal{T}_\infty$, $\mathcal{S}$ and $\mathcal{N}$. (N.B., φ does not in general commute with R_∞.) The following congruences are enjoyed by $\mathcal{N}$.

Lemma 13. *Let $g \in 1+\pi^i I$, $i \geqq 1$, and $h \in \mathcal{M}$. Then*

(i) $\mathcal{N}(g) \equiv 1 \bmod \pi^{i+1} I$,

(ii) $\dfrac{\mathcal{N}^i(h)}{\varphi \mathcal{N}^{i-1}(h)} \equiv 1 \bmod \pi^i I$.

Proof. We first observe that ${}_u g \equiv g \bmod \pi^i \mathfrak{p}_0$, for $u \in \mathfrak{F}_0$. Hence $\mathcal{Q}(g) \equiv g^q \equiv 1 \bmod \pi^{i+1}$ since $\mathcal{Q}(g) \in I$. We deduce from (1) that $\mathcal{N}(g)_\pi \equiv 1 \bmod \pi^{i+1}$. To conclude the proof of (i), we must verify the statement: If $k \in I$ and $k_\pi \in \pi^n I$ then $k \in \pi^n I$. We proceed by induction on n. The statement is trivial for $n=0$. Suppose now $n>0$ and the statement is true for $n-1$. Set $k'=\pi^{1-n}k$. Then $k'_\pi \in \pi I$ and by induction $k' \in I$. As $[\pi](T) \equiv T^q \bmod \pi$ we have $k'(T^q) \in \pi I$, and so $k=\pi^{n-1}k' \in \pi^n I$. Thus we have (i)

We first prove (ii) for $i=1$. Suppose $f \in I$, then as $\mathcal{N}(f)_\pi (T) \equiv \mathcal{N}(f)(T^q) \bmod \pi$ and ${}_v f \equiv f \bmod \mathfrak{p}_0$, we may conclude from (1) that $\mathcal{N}(f)(T^q) \equiv f(T)^q \bmod \pi I$. But, $f(T)^q \equiv \varphi f(T^q) \bmod \pi I$; we deduce

$$\mathcal{N}(f) \equiv \varphi f \bmod \pi I, \tag{2}$$

for $f \in I$. It follows from (2) that $\mathcal{N}(f)/\varphi f \equiv 1 \bmod (\varphi f)^{-1} \cdot \pi I$, for $f \in \mathcal{M} \cap I$. But, by Corollary 12 (i) $v_T \mathcal{N}(f) = v_T f$ so that we actually have

$$\mathcal{N}(f)/\varphi f \equiv 1 \bmod \pi I, \tag{3}$$

for $f \in \mathcal{M} \cap I$. Since, for every $f \in \mathcal{M}$, either f or $f^{-1} \in I$, (3) follows immediately for any $f \in \mathcal{M}$. We obtain (ii) by applying (i) iteratively to (3).

It is an immediate consequence of this lemma that the limit,

$$\mathcal{N}^\infty(f) \overset{\text{defn.}}{=} \lim_{i \to \infty} \varphi^{-i} \mathcal{N}^i(f)$$

exists for all $f \in \mathcal{M}$ and satisfies

$$\mathcal{N}(\mathcal{N}^\infty(f)) = \varphi \mathcal{N}^\infty(f) \quad \text{and} \quad \mathcal{N}^\infty(f)/f \equiv 1 \bmod \pi I \tag{4}$$

Example. Let $H=K=\mathbb{Q}_p$ be the field of p-adic numbers. Let $\mathfrak{F}=\mathbb{G}_m$ be the formal multiplicative group. Then $\mathbb{G}_m$ is a Lubin-Tate group over $\mathbb{Q}_p$ corresponding to the parameter p. We then have

$$\mathcal{N}^\infty(1-T)=(\varepsilon-1)-T, \qquad \mathcal{N}^\infty(1-(1+T)^a)=1-(1+T)^a$$

where ε is the unique element of $\mathbb{Z}_p$ which satisfies $\varepsilon^p=\varepsilon$ and $\varepsilon \equiv 2 \bmod p$, and where $(a,p)=1$. Evaluating $(1-(1+T)^a)/T$ at $u_n=\zeta_n-1$ $(\zeta_n^{p^{n+1}}=1)$ we obtain of course the circular units.

Now set $\mathcal{M}_\infty = \{f \in \mathcal{M} / \mathcal{N}(f) = \varphi f\}$ and $\mathcal{M}^o_\infty = \mathcal{M}^o \cap \mathcal{M}_\infty$. We see from (4) that $\mathcal{M}_\infty$ is non-trivial and that $\mathcal{N}^\infty$ is a projector from $\mathcal{M}$ onto $\mathcal{M}_\infty$; in fact we have:

Proposition 14. (i) $\mathcal{M}_\infty = (\mathcal{N}^\infty(T))^{\mathbb{Z}} \times V \times \mathcal{M}^o_\infty$. (ii) *The sequence:*

$$(1) \to 1 + \pi I \to \mathcal{M}^o \underset{\iota}{\overset{\mathcal{N}^\infty}{\rightleftarrows}} \mathcal{M}^o_\infty \to (1)$$

is a split exact sequence of topological $\mathcal{T}_\infty$-modulues, where ι is the inclusion.

Proof. This follows immediately from Corollary 12 and Lemma 13.

Let $G(T) = \mathcal{N}^\infty(T)$. We then have $G \in T \cdot \mathcal{M}^o$.
We now prove the main result of this section.

Theorem 15. *Let $\alpha \in H_n^*$, then there exists an $f_\alpha \in \mathcal{M}$ such that*

$$\varepsilon_i f_\alpha = \varphi^i N_{n,i} \alpha \qquad \text{for } 0 \leqq i \leqq n.$$

Proof. For all $\beta \in H_n$, let $h_\beta \in \mathcal{H}$ (cf. Sect. III) be defined as follows:

$$h_\beta(\sigma v_i) = \begin{cases} 0 & i > n, \\ \sigma \varphi^i N_{n,i} \beta & 0 \leqq i \leqq n, \end{cases} \qquad \text{for } \sigma \in G_\infty.$$

First suppose $\alpha \in \mathcal{O}_n^*$. We will show $h_\alpha \in \mathcal{H}_n$, i.e., we will show

$$\int_{\mathfrak{F}} g \cdot h_\alpha \equiv 0 \bmod \pi^{n+1} \tag{5}$$

for all $g \in T \cdot I$. It is sufficient to verify (5) for all $g \in \{G(T)^k\}_{k \geqq 1}$. Indeed, the collection of $\mathcal{O}_H$-linear combinations of elements in this set is dense in $T \cdot I$, and (5) is clearly a continuous $\mathcal{O}_H$-linear condition on g. But, in view of the definition of $G(T)$ and Corollary 12 (ii)

$$G(T)^k \cdot h_\alpha = h_\gamma$$

where $\gamma = \varphi^{-n} G(u_n)^k \cdot \alpha$. As $\gamma \in \mathfrak{p}_n$, for $k \geqq 1$, we are reduced to proving

$$\int_{\mathfrak{F}} h_\beta \equiv 0 \bmod \pi^{n+1} \tag{6}$$

for all $\beta \in \mathfrak{p}_n$.

Now choose $f \in T \cdot I$ so that $\varepsilon_n f = \varphi^n \beta$. From Corollary 5 (ii), we deduce that

$$T_i(\varepsilon_i g) = (\mathcal{S}^{i+1}(g) - \mathcal{S}^i(g))(0),$$

for $g \in I$. Let $\mathcal{N}^{(i)}(f) = \varphi^{-i} \mathcal{N}^i(f)$, for $j \geqq 0$. Using (III, 3), Corollary 12 (ii) and Corollary 5 (i) we have

$$\begin{aligned}\int_{\mathfrak{F}} h_\beta &= \sum_{i=0}^{n} T_i(\varphi^i N_{n,i}\beta) \\ &= \sum_{i=0}^{n} (\mathcal{S}^{i+1}(\mathcal{N}^{(n-i)}(f)) - \mathcal{S}^i(\mathcal{N}^{(1-i)}(f)))(0) \\ &= \mathcal{S}^{n+1}(f)(0) - \mathcal{N}^{(n)}(f)(0) + \sum_{i=0}^{n} \mathcal{S}^{i+1}(\mathcal{N}^{(n-(i-1))}(f) - \mathcal{N}^{(n-i)}(f))(0).\end{aligned}$$

Now, $\mathcal{S}^{n+1}(f) \equiv 0 \bmod \pi^{n+1}$ by Lemma 6, $\mathcal{N}^{(n)}(f)(0)=0$ since $1 \leqq v_T(f) = v_T(\mathcal{N}^{(n)}(f))$, by Corollary 12 (i), and

$$\mathcal{S}^{i+1}(\mathcal{N}^{(n-(i-1))}(f) - \mathcal{N}^{(n-i)}(f)) \equiv 0 \bmod \pi^{n+1} \cdot I$$

by Lemma 6 (ii) and Lemma 13 (ii). Hence we have (6) and so $h_\alpha \in \mathcal{H}_n$ for all $\alpha \in \mathcal{O}_n^*$. Thus by Theorem 8 there exists an $f_\alpha \in I$ such that $f_\alpha(u) = h_\alpha(u)$ for all $u \in \mathfrak{F}_n$ Actually, $f_\alpha \in I^*$ since $f_\alpha(0) \equiv \alpha \bmod \mathfrak{p}_n$. Thus we have Theorem 15 for $\alpha \in \mathcal{O}_n^*$.

It remains to consider β arbitrary in H_n^*. But, $H_n^* = (\varphi^{-n} G(v_n))^{\mathbb{Z}} \times \mathcal{O}_n^*$, as $G(T) \in T \cdot \mathcal{M}^o$. So if $\beta = (\varphi^{-n} G(v_n))^k \cdot \alpha$, where $k \in \mathbb{Z}$ and $\alpha \in \mathcal{O}_n^*$ we set

$$f_\beta = G(T)^k \cdot f_\alpha$$

where f_α is as above. This completes the proof of the theorem.

Let $X_\infty = \varprojlim H_n^*$ and $X_\infty^0 = \varprojlim 1 + \mathfrak{p}_n$, where the inverse limits are taken with respect to the maps, $N_{m,n}$ for $m \geqq n \geqq 0$. $X_\infty^o \subseteq X_\infty$ and X_∞^o is naturally a $\mathcal{T}_\infty$-module.

Theorem 16. *There exists a unique map $\gamma (= \gamma_v)$ from X_∞ into $\mathcal{M}$ such that*

$$\begin{tikzcd} X_\infty \arrow[r, "\gamma"] \arrow[d, "N_{\infty,\iota}"] & \mathcal{M} \arrow[d, "\varepsilon_\iota"] \\ H_j^* \arrow[r, "\varphi^\iota"] & H_i^* \end{tikzcd} \tag{7}$$

commutes for all $i \geqq 0$.

Proof. The uniqueness of γ follows from our uniqueness principle.

We know from the previous theorem that for each positive integer n there exists a (non-unique) map $\gamma_n \colon X_\infty \to \mathcal{M}$, which makes (7) commute for all i such that $0 \leqq i \leqq n$. Also, it is easy to see that $v_{\mathfrak{p}_n}(b_n) = v_T(\gamma_n(b))$ and $v_{u_m}(b_m) = v_{\mathfrak{p}_n}(b_n)$ for all $b \in X_\infty$ and all $m \geqq n$ (H_∞/H is totally ramified). Thus for each $b \in X_\infty$ the sequence $\{\gamma_n(b)\}$ is contained in $T^k \mathcal{O}_H[[T]]$, for some $k \in \mathbb{Z}$. Lemma 2a implies that this sequence is Cauchy and if we set $\gamma(b)$ equal to its limit for each $b \in X_\infty$, we see immediately that γ makes (1) commute, for all $n \geqq 0$. This completes the proof of the Theorem.

Corollary 17. *γ is a topological isomorphism from X_∞ onto $\mathcal{M}_\infty$, which restricts to a topological $\mathcal{T}_\infty$-isomorphism from X_∞^o onto $\mathcal{M}_\infty^o$.*

Proof. As $\varepsilon_n \mathcal{N}(\gamma(b)) = N_{n+1,n}(\varepsilon_{n+1}\gamma(b)) = \varphi\varepsilon_n\gamma(b) = \varepsilon_n\varphi\gamma(b)$ for all $n \geqq 0$, it follows from our uniqueness principle that $\mathcal{N}\gamma(b) = \varphi\gamma(b)$ so that $\gamma(b) \in \mathcal{M}_\infty$ for all $b \in X_\infty$. Also, given $f \in \mathcal{M}_\infty$, the sequence $c_f = \{\varphi^{-1}\varepsilon_i(f)\}$ lies in X_∞. Moreover, the map $f \mapsto c_f$ is clearly the inverse of γ on $\mathcal{M}_\infty$ and is continuous. Hence, γ is a bijection from X_∞ to $\mathcal{M}_\infty$.

Obviously γ is a homomorphism, and by Lemma 2a, it is continuous. This together with the continuity of γ^{-1} implies that γ is a topological isomorphism onto $\mathcal{M}_\infty$.

Finally, it is plain that γ restricted to X_∞^o is a $\mathbb{Z}[G_\infty]$-homomorphism onto $\mathcal{M}_\infty^o$, so as γ is continuous, it follows that γ defines a topological $\mathcal{T}_\infty$-isomorphism from X_∞^o onto $\mathcal{M}_\infty^o$.

Corollary 18. $X_\infty \approx \mathfrak{k}((T))^*$, *where* $\mathfrak{k}$ *is the residue field of* H.

Proof. This follows immediately from the previous theorem and Proposition 14.

Remark. By means of this isomorphism, one can give $X_\infty \cup \{0\}$ the structure of a field. It can be shown that this is identical with Fontaine's construction [F], by means of which one can set up a correspondence between algebraic extensions of H_∞ and algebraic extensions of $\mathfrak{k}((T))$.

Corollary 19. *If* $[H:K] < \infty$ *then* $(\bigcap_{m \geqq n} N_{m,n}(1+\mathfrak{p}_m)) \cdot (1+\pi\mathcal{O}_n) = 1+\mathfrak{p}_n$, *for* $n \geqq 0$.

Proof. This follows immediately from the preceding Theorem, Proposition 14 and the compactness of $1+\mathfrak{p}_n$.

V. The Lubin-Tate Logarithm

We maintain the notations of the previous section. In this section we will study the logarithm associated with $\mathfrak{F}$. We begin with a technical lemma on the endomorphism $[\pi]$. Let $\mathfrak{p}_H$ denote the ideal $\pi\mathcal{O}_H$.

Lemma 20. *The following assertions are true for each integer* $i \geqq 1$. (i) *If* g, h *are power series in* I, *with* g *arbitrary and* h *satisfying* $h(T) \equiv T^q \bmod \mathfrak{p}_H$, *then* $[\pi^i] \circ g \equiv [\pi^{i-1}] \circ \varphi g \circ h \bmod \mathfrak{p}_H^i$; (ii) *Let* N *be a positive integer, and* r, ε *positive reals, with* $r<1$. *If* f *belongs to* $S_H(r,\varepsilon) \cap T^N H[[T]]$, *then* $f_{\pi^i} \in S_H(r, \varepsilon q^{-iN} M_r^N)$, *where* M_r *is a positive real number depending only on* r.

Proof. The proof of (i) is by induction on i. It is true for $i=1$, because

$$[\pi] \circ g \equiv g^q \equiv \varphi g(T^q) \equiv \varphi g \circ h \bmod \mathfrak{p}_H.$$

Suppose that it is true for an integer $i \geqq 1$. Write

$$[\pi^i] \circ g = [\pi^{i-1}] \circ \varphi g \circ h + \pi^i k_1, \qquad [\pi](T) = T^q + \pi k_2.$$

where k_1, k_2 belong to I. Then, if we put $A = [\pi^{i-1}] \circ \varphi g \circ h$, we have

$$\begin{aligned}[\pi^{i+1}]\circ g &= (A+\pi^i k_1)^q+\pi k_2(A+\pi^i k_1)\\ &\equiv A^q+\pi k_2(A) \bmod \mathfrak{p}_H^{i+1},\\ &\equiv [\pi^i]\circ \varphi g\circ h \bmod \mathfrak{p}_H^{i+1},\end{aligned}$$

and so (i) follows.

To prove (ii) we first observe that if $h\in B$, then

$$|[\pi](b)| \leqq \max(q^{-1}|b|, |b|^q),$$

since $[\pi](T)=\pi T+T^q+\pi g(T)$, for some $g\in T^2\mathcal{O}_K[[T]]$. From this it follows that there exists a positive integer k_r, depending only on $r=|b|$, such that $|[\pi^{k_r}](b)|<q^{-(q-1)^{-1}}$. It also follows that $|[\pi](b)|<q^{-1}|b|$ if $|b|<q^{-(q-1)^{-1}}$. From these two facts we see that there exists a constant $C_r>0$, depending only on r, such that

$$|[\pi^i](b)|<q^{-i}C_r$$

if $|b|=r$. Now take f to be a power series in $S_H(r,\varepsilon)\cap T^N H[[T]]$, say $f(T) = T^N h(T)$. It is clear from the above that

$$|f_{\pi^i}(b)|<q^{-iN}C_r^N|h_{\pi^i}(b)|.$$

But $|[\pi^i](b)|<r$, and so if $r>0$,

$$|h([\pi^i](b))|<\sup_{|x|=r}\left|\frac{f(x)}{x^N}\right|<r^{-N}\varepsilon.$$

Combining these last two inequalities, assertion (ii) follows with $M_r=C_r/r$. This completes the proof of Lemma 20.

Let $\lambda=\lambda_{\mathfrak{F}}$ denote the logarithm of $\mathfrak{F}$. Thus, if $\mathbb{G}_a$ denotes the formal additive group, then λ is the unique power series in $K[[T]]$ which, gives an isomorphism

$$\lambda\colon \mathfrak{F} \xrightarrow{\sim} \mathbb{G}_a$$

over K, and which satisfies $\lambda(T)\equiv T \bmod$ degree 2. We have the following basic lemma about λ.

Lemma 21. (i) $\lambda=\lim_{n\to\infty}\pi^{-n}[\pi^n]$, *where the limit is taken in* $K((T))_1$; (ii) *If* g *and* h *satisfy the same hypotheses as in part* (i) *of Lemma* 20, *then,*

$$\lambda\circ g-\frac{\lambda\circ\varphi g\circ h}{\pi}\in I.$$

Proof. Let $K[[T]]_1=K[[T]]\cap K((T))_1$, endowed with the restriction topology. Put $h_n=\pi^{-n}[\pi^n]$. Since $K[[T]]_1$ is a complete, non-archimedean, topological ring, it suffices to show that $h_{n+1}-h_n$ tends to 0 as $n\to\infty$, to prove the existence of $\lim_{n\to\infty} h_n$. Put $g(T)=h_1(T)-T$. Obvously, $g(T)$ is divisible by T^2 in $K[[T]]_1$, and

$$\pi^{-n}g\circ[\pi^n]=h_{n+1}-h_n.$$

Applying (ii) of Lemma 20 with $N=2$, we conclude that the left hand side of this equation tends to zero as $n\to\infty$. Thus $h=\lim_{n\to\infty} h_n$ exists. Since $h(T)\equiv T \bmod$ degree 2, and $h\circ[\pi]=\pi h$, a simple argument based on the uniqueness of λ, which we omit, shows that we must have $h=\lambda$. This completes the proof of (i). Also (i) implies that

$$\lambda\circ g-\frac{\lambda\circ\varphi g\circ h}{\pi}=\lim_{n\to\infty}\left(\frac{[\pi^{n+1}]\circ g-[\pi^n]\circ\varphi g\circ h}{\pi^{n+1}}\right),$$

and the right hand side is in I by part (i) of Lemma 20. This completes the proof of Lemma 21.

For a more detailed discussion of λ, see [Fr] Chapt. 4 (also see [W]). Here we simply recall, without proof, the following basic facts. First, λ commutes with the action of $\mathcal{O}_K$ as endomorphism ring of $\mathfrak{F}$ and $\mathbb{G}_a$. Second, $\frac{d}{dT}\lambda$ belongs to $\mathcal{O}_K[[T]]$. Third, λ converges on B and if $|b|<q^{-(q-1)^{-1}}$, then

$$|\lambda(b)|=|b|. \tag{1}$$

Finally, the kernel of λ on B is the group $\mathfrak{F}_\infty$. We point out, however, that all of these facts follow easily from Lemma 21.

If $\mathscr{I}$ is any topologically nilpotent ideal in a topological $\mathcal{O}_K$-algebra R, we define $\mathfrak{F}(\mathscr{I})$ to be the topological group whose underlying topological space is $\mathscr{I}$ and whose addition law "$[+]$" is given by

$$a[+]b=\mathfrak{F}(a,b).$$

(N.B. This is consistent with our earlier use of $[+]$.)

Let $\mathscr{A}_\infty$ denote the closure of $\mathcal{O}_K[G_\infty]$ in R_∞. Let $\mathscr{m}$ denote the maximal ideal of I. We may give the group $\mathfrak{F}(\mathscr{m})$ the structure of a continuous $\mathscr{A}_\infty$-module such that the action of $\omega\in\mathscr{A}_\infty$ on an element f of $\mathfrak{F}(\mathscr{m})$ is denoted by $[\omega](f)$, and such that

$$[a](f)=[a]\circ f,\qquad [\sigma](f)=f\circ[\kappa(\sigma)],$$

for $a\in\mathcal{O}_K$, $\sigma\in G_\infty$. Once again, the proof of the existence and uniqueness of such a structure runs along similar lines to the proof of Theorem 1.

We shall henceforth always consider $\mathfrak{F}(\mathscr{m})$ endowed with this $\mathscr{A}_\infty$-module structure. It is plain that λ defines a continuous $\mathscr{A}_\infty$-module homomorphism from $\mathfrak{F}(\mathscr{m})$ into $H[[T]]_1$. The remainder of this section will be devoted to giving descriptions of $\mathfrak{F}(\mathscr{m})$ and of $\lambda(\mathfrak{F}(\mathscr{m}))$. Let $\Theta\colon H[[T]]_1\to H[[T]]_1$ be the map defined by

$$\Theta(f)=f-\frac{\varphi f_\pi}{\pi}.$$

Let $\Theta_{\mathfrak{F}}\colon \mathfrak{F}(\mathscr{m})\to \mathrm{H}[[T]]_1$ be the map defined by

$$\Theta_{\mathfrak{F}}(g)=\Theta(\lambda(g)).$$

Let

$$A = \left\{ g \in I : g(0) \in \mathscr{J}_H, \frac{d}{dT}(g)(0) \in (1-\varphi)\, \mathcal{O}_H \right\},$$

where $\mathscr{J}_H = (1-\varphi)\,\mathcal{O}_H + \mathfrak{p} H$, if $q=2$, and $\mathscr{J}_H = \mathcal{O}_H$ otherwise. Finally let $\mathscr{C} = \mathfrak{F}_\infty \cap \mathfrak{F}(\mathfrak{p}_H)$. Since $[H_0 : H] = q-1$, $\mathscr{C} = \{0\}$ unless $q=2$, in which case $\mathscr{C} = \mathfrak{F}_\pi$. We may now state:

Theorem 22. *The sequence*

$$0 \to \mathscr{C}[+][\mathscr{A}_\infty]\, T \xrightarrow{\text{incl.}} \mathfrak{F}(\mathscr{m}) \xrightarrow{\Theta_{\mathfrak{F}}} A \to 0$$

is an exact sequence of topological $\mathscr{A}_\infty$-modules, which splits when $q \neq 2$ or $H = K$.

Proof. The fact that $\Theta_{\mathfrak{F}}$ is a continuous $\mathscr{A}_\infty$-homomorphism follows from the fact that both λ and Θ obviously are.

We see that the image of $\Theta_{\mathfrak{F}}$ is contained in I from Lemma 2 (ii) and the fact that $\pi^{-1}\mathfrak{p}_H = \mathcal{O}_H$. To see that its image is A, we first observe:

$$\Theta_{\mathfrak{F}}(f)(0) = \lambda(f(0)) - \frac{\varphi\,\lambda(f(0))}{\pi}. \tag{2}$$

By (1), $|\lambda(a)| = |a|$ if $a \in \mathfrak{p}_H$ and $q \neq 2$ or $a \in \mathfrak{p}_H^2$ and $q=2$; since $|\pi^i| = q^{-i} > q^{-(q-1)^{-1}}$ if $i \geqq 1$ and $q \neq 2$ or $i \geqq 2$ and $q=2$. Furthermore, if $q = 2 H = K$, $\dfrac{\lambda(\pi a)}{\pi} \equiv (1-\varphi)\, a \bmod \mathfrak{p}_H$. Thus $\lambda(\mathfrak{p}_H) = (1-\varphi)\,\mathfrak{p}_H + \mathfrak{p}_H^2$. Therefore the right hand side of (2) is in $\mathscr{J}_H$ and, in fact, $\Theta_{\mathfrak{F}}(\mathfrak{p}_H) = \mathscr{J}_H$. Second, we see easily

$$\left(\frac{d}{dT}\,\Theta_{\mathfrak{F}}(f)\right)(0) = (1-\varphi)\left(\frac{d}{dT}\,\lambda(f)\right)(0).$$

As $\lambda' \in \mathcal{O}_K[[T]]$, the right hand side of the above expression lies in $(1-\varphi)\mathcal{O}_H$ for all $f \in \mathfrak{F}(\mathscr{m})$ and this together with the above shows that the image of $\Theta_{\mathfrak{F}} \subseteq A$. To see the other inclusion we set $f_{\varepsilon,i} = \Theta_{\mathfrak{F}}(\varepsilon T^i)$ for each integer $i \geqq 1$ and each $\varepsilon \in V$, where V is the set of roots of unity in H of order prime to p. We see that

$f_{\varepsilon,i}(T) \equiv (\varepsilon - \varepsilon^q \pi^{i-1})\, T^i \bmod$ degree $i+1$. It follows from this that if $h \in A$ then there exist $a_{\varepsilon,i} \in \mathcal{O}_K$ and an $a \in \mathscr{J}_H$ such that

$$h = a + \sum_{v,i} a_{\varepsilon,i} \cdot f_{\varepsilon,i}.$$

It follows that for each i, $|a_{\varepsilon,i}| < \delta$ for all but finitely many ε where δ is any possitive real. From this and Lemma 20 (ii) it follows that the series $g = \sum_{v,i}{}_{\mathfrak{F}}\, [a_{v,i}](v T^i)$ converges in $\mathfrak{F}(\mathscr{m})$, where $\Sigma_{\mathfrak{F}}$ denotes summation in $\mathfrak{F}(\mathscr{m})$. From the continuity of $\Theta_{\mathfrak{F}}$ we see that $h - \Theta_{\mathfrak{F}}(g) = a$. Since $\Theta_{\mathfrak{F}}(\mathfrak{p}_H) = \mathscr{J}_H$ we deduce that $\Theta_{\mathfrak{F}}(\mathfrak{F}(\mathscr{m})) = A$.

Now suppose $\Theta_{\mathfrak{F}}(g) = 0$. Let $f = \lambda(g)$; then $\Theta(f) = 0$. Suppose $f \neq 0$. If $a T^k$ is the first non-vanishing term of $f(T)$ we conclude, by examining the coefficient of T^k in $\Theta(f)$ as above, that $k=1$ and $a \in K$. But, then $f - a\lambda$ satisfies the same conditions as f, has no linear term and therefore must be zero. Hence we conclude, $\Theta_{\mathfrak{F}}(g)=0$ if and only if $\lambda(g)=a\lambda$ for some $a\in K$. Since the kernel of λ on $\mathfrak{F}(m)$ is $\mathscr{C}$, we have immediately that $g\in\mathscr{C}[+][\mathscr{A}_\infty]\,T$, as asserted.

As for the splitting, first observe that $[\sigma]T=[\kappa(\sigma)](T)$, $\sigma\in G_\infty$, so that $[\mathscr{A}_\infty]_T=[\mathcal{O}_K]_T$. We define $P\colon \mathfrak{F}(m)\to\mathscr{C}[+][\mathscr{A}_\infty]_T$ by setting $P(f)=[f(0)](\alpha)[+]\left[\frac{d}{dT}(f)(0)\right](T)$, where α is the generator of $\mathscr{C}$ (recall $|\mathscr{C}|\leqq 2$). It is easy to see that P is an $\mathcal{O}_K[G_\infty]$ homomorphism, hence as P is clearly continuous, P is an $\mathscr{A}_\infty$-homomorphism. It is also plain that P is a projector onto the kernel of $\Theta_{\mathfrak{F}}$.

To see that the sequence is topologically exact, we first observe that all the maps are continuous and that the first group is compact. Therefore, we only need to show that $\Theta_{\mathfrak{F}}$ is an open mapping. But this follows immediately from the facts that $\Theta_{\mathfrak{F}}(\mathfrak{F}(\pi^n m))\supseteq\pi^{n+1}A$ and that $\Theta_{\mathfrak{F}}(\mathfrak{F}(T^k\cdot\mathcal{O}_H[[T]]))=T^k\mathcal{O}_H[[T]]$ for $n\geqq 1$ and $k\geqq 2$. These facts, in turn, may be proven by arguments similar to those given above. This completes the proof of the Theorem.

Corollary 23. *When $q\neq 2$ or $H=K$, $\mathfrak{F}(m)$ is $\mathscr{A}_\infty$-isomorphic to $\mathscr{C}\oplus\mathscr{A}_\infty/\mathscr{I}_1\oplus A$, where $\mathscr{I}_1$ is the closed ideal in $\mathscr{A}_\infty$ generated topologically by $\{1-\kappa(\sigma)^{-1}\sigma\}_{\sigma\in G_\infty}$.*

Proof. From the previous theorem we see that it is sufficient to show that the kernel of the map $\omega\mapsto[\omega]T$ from $\mathscr{A}_\infty$ to $\mathfrak{F}(m)$ is precisely $\mathscr{I}_1$. Because $[\sigma]T=[\kappa(\sigma)](T)$ we see that there is a unique map $\mathscr{K}_1\colon \mathscr{A}_\infty\to\mathcal{O}_K$ such that $[\omega]T=[\mathscr{K}_1(\omega)]\,T$. It is also clear that the map $\mathscr{K}_1$ is a continuous surjective ring homomorphism whose kernel is identical with that of $\omega\mapsto[\omega]\,T$. We see easily that $\mathscr{I}$, is contained in the kernel of $\mathscr{K}_1$. Now $\mathscr{I}_1+\mathcal{O}_K=\mathscr{A}_\infty$ since $\mathscr{I}_1+\mathcal{O}_K$ is closed and contains $\mathcal{O}_K[G_\infty]$; hence as $\mathscr{K}_1(a)=\mathscr{A}$ for $a\in\mathcal{O}_K\subseteq\mathscr{A}_\infty$ it follows that $\mathscr{I}_1$ is the kernel of $\mathscr{K}_1$ and our proof is complete.

Now let $\Xi\colon T^2\mathcal{O}_H[[T]]\to T^2H[[T]]_1$ be defined by

$$\Xi f=\sum_{i=0}^{\infty}\frac{\varphi^i f_{\pi^i}}{\pi^i}$$

That Ξ is a well defined continuous map follows from Lemma 20 (ii). It is also plain that Ξ is an $\mathscr{A}_\infty$ homomorphism. For $b\in\mathcal{O}_H$ set

$$\rho(b)=b\lambda+\sum_{i=0}^{\infty}\varphi^i a\left(\frac{[\pi^i]}{\pi^i}-\lambda\right),$$

where $a=(1-\varphi)b$. The series converges by Lemma 20 (ii) since $[\pi^i]-\pi^i\lambda=(T-\lambda)\circ[\pi^i]$.

Theorem 24. *Let $h\in H[[T]]_1$. Then $h\in\lambda\mathfrak{F}(m)$ if and only if there exists elements $a\in\pi\mathscr{J}_H$, $b\in\mathcal{O}_H$ and $f\in T^2\cdot\mathcal{O}_H[[T]]$ such that*

$$h=a+\rho(b)+\Xi(f).$$

Moreover, $\lambda(\mathfrak{F}(T^k\cdot\mathcal{O}_H[[T]]))=\Xi(T^k\cdot\mathcal{O}_H[[T]])$, for $k\geqq 2$.

Proof. We see easily that

$$\Theta\,\Xi(f)=f,\qquad \Theta\,\rho(b)=(1-\varphi)\,b\,T\quad\text{and}\quad \Theta(\pi\mathscr{J}_H)=\mathscr{J}_H, \tag{3}$$

where $f \in T^2 \cdot I$, $b \in \mathcal{O}_H$. Let $\mathscr{E}$ denote the set of all elements of the form $a+\rho(b)+\Xi(f)$, where a, b and f are as above. It follows from (3) that $\Theta\mathscr{E}=A$. Thus by Theorem 22, $\mathscr{E}+\operatorname{Ker}\Theta=\lambda(\mathfrak{F}(\mathscr{m}))+\operatorname{Ker}\Theta$. Therefore since $\mathscr{E}\supseteq\{\rho(b)\}_{b\in\mathcal{O}_K}=\mathcal{O}_K\cdot\lambda(\mathfrak{F}(\mathscr{m}))\cap\operatorname{Ker}\Theta$, to prove the first part of the theorem, we need only show that $\mathscr{E}\subseteq\lambda(\mathfrak{F}(\mathscr{m}))$. We know from the proof of Theorem 22 that $\operatorname{Ker}\Theta=K\cdot\lambda$. Let $g=a+\rho(b)+\Xi(f)$ be an element of $\mathscr{E}$. Suppose $g=\lambda(h)+c\lambda$ with $h\in\mathscr{m}$ and $c\in K$. We then see that

$$b=\frac{d}{dT}\rho(b)(0)=\frac{d}{dT}(g)(0)=c\lambda'(0)+\lambda'(h)(0)\cdot h'(0).$$

So as $\lambda'\in\mathcal{O}_K[[T]]$ and $\lambda'(0)=1$ we see that $b\equiv c \bmod \mathcal{O}_H$. But, $b\in\mathcal{O}_H$ so that $c\in\mathcal{O}_H\cap K=\mathcal{O}_K$ and $g=\lambda([c](T)[+]h)$. Thus $\mathscr{E}\subseteq\lambda(\mathfrak{F}(\mathscr{m}))$.

As for the second assertion, it follows from (4), the fact that $\lambda(T^k I)\subseteq T^k H[[T]]_1$, and the fact that Θ is an injection on $T^k H[[T]]_1$, for $k\geqq 2$.

Corollary 25 (Iwasawa). *If $\mathfrak{p}_n$ is the maximal ideal of the ring of integers in $\mathbb{Q}_p(\mathfrak{p}_n)$ where ζ_n is a primitive p^{n+1}-st root of unity, then*

$$(1-a^{-1}\sigma(a))\,\gamma_n\in\operatorname{Log}(1+\mathfrak{p}_n)$$

for $a\in\mathbb{Z}_p^*$, where $\sigma(a)$ is the element of Gal $(\mathbb{Q}_p(\zeta_n)/\mathbb{Q}_p)$ such that $\sigma(a)\,\zeta_n=\zeta_n^a$, and where

$$\gamma_n=\sum_{i=0}^{n}\frac{\zeta_n^{p^i}-1}{p^i}.$$

Proof. We first note that if we set $H=K=\mathbb{Q}_p$, $\pi=p$ and $\mathfrak{F}=\mathbb{G}_m$ then $H_n=\mathbb{Q}_p(\zeta_n)$. Now let $\sigma=\kappa^{-1}(a)$ and set

$$g(T)=(1-a^{-1}\sigma)\,T=T-a^{-1}((1+T)^a-1).$$

Clearly $g\in T^2\mathbb{Z}_p[[T]]$; and so by Lemma 20 (ii) we may set

$$f(T)=\Xi g=\sum_{i=0}^{\infty}\frac{g([p^i])}{p^i}\qquad([p](T)=(T+1)^p-1).$$

Thus by the above Theorem, $f(T)=\operatorname{Log}(1+h(T))$ for some h in $T^2\mathbb{Z}_p[[T]]$. Therefore

$$f(\zeta_n-1)=\operatorname{Log}(1+h(\zeta_n-1))\in\operatorname{Log}(1+\mathfrak{p}_n).$$

But it is plain that

$$f(\zeta_n-1)=(1-a^{-1}\sigma(a))\,\gamma_n.$$

This completes the proof of the corollary.

Remark. Iwasama's original proof of this fact [I_1] involved the explicit reciprocity laws. He went on the show, by means of an index computation, that for p odd,

$$\operatorname{Log}(1+\mathfrak{p}_n)=p\mathbb{Z}_p+\mathscr{I}_1\gamma_n,$$

where $\mathscr{I}_1$ is as in Corollary 23. In another paper, we will show how this result fits into a general picture and can be extended to the case of arbitrary division towers of Lubin-Tate formal groups of height one.

VI. The Dual of the Image of the Logarithm

In this section H shall denote a *finite* unramified extension of K. Since G_n acts on $\mathfrak{p}_n$ we may make $\mathfrak{F}(\mathfrak{p}_n)$ into an $\mathcal{O}_K[G_n]$-module and hence a continuous $\mathscr{A}_\infty$-module in the natural way. It is clear that λ then defines an $\mathscr{A}_\infty$-module homomorphism from $\mathfrak{F}(\mathfrak{p}_n)$ into H_n^+.

In order to study the image of $\mathfrak{F}(\mathfrak{p}_n)$ under λ, and its dual, we introduce for each $n\geqq 0$, a continuous $\mathscr{A}_\infty$-homomorphism L_n: $TH[[T]]_1\to H_n$ defined by

$$L_n(f)=\sum_{i=0}^{n}\frac{\varphi^i\varepsilon_{n-i}(f)}{\pi^i}$$

We have immediately

$$L_n(f)=\varepsilon_n(f)+\pi^{-1}\varphi L_{n-1}(f) \tag{1}$$

for $n\geqq 1$, and also

$$L_n(\Theta(f))=\varepsilon_n f,$$

where Θ is as in the previous section. From this and Theorem 22, we deduce

$$L_n(A')=\varepsilon_n\lambda(T\cdot I)=\lambda(\mathfrak{p}_n)\overset{\text{defn.}}{=}\mathscr{L}_n, \tag{2}$$

where $A'=\{f\in A/f(0)=0\}$.

Recall that $T_{n/K}$ denotes the trace from H_n to K. We define the trace pairing

$$\langle\ ,\ \rangle_n\colon H_n\times H_n\to K$$

by setting $\langle a,b\rangle_n=T_{n/K}(ab)$. Thus $\langle\ ,\ \rangle_n$ is a symmetric, non-degenerate, K-bilinear pairing. We define $\mathfrak{X}_n=\{a\in H_n/\langle a,b\rangle_n\in\mathcal{O}_K$ for all $b\in\mathscr{L}_n\}$. Then $\mathfrak{X}_n$ is a compact $\mathscr{A}_\infty$-submodule of $H_n^+(\mathscr{L}_n\supseteq\mathfrak{p}_H^2\mathcal{O}_n$ by $(V,1))$.[1] We now prove:

Theorem 26. *If $\alpha\in H_n$ then $\alpha\in\mathfrak{X}_n$ if and only if there exists an element $f_\alpha\in T^{-1}\mathcal{O}_H[[T]]$ such that $\operatorname{Res}_0 f_\alpha\in\mathcal{O}_K$ and*

$$\varepsilon_i f_\alpha=\pi^{i+1}\varphi^i T_{n,i}(\alpha), \tag{3}$$

for $0\leqq i\leqq n$.

[1] In the following we let $\operatorname{Res}_0 f$ denote the coefficient of T^{-1} in f where $f\in H_\infty((T))$

Proof. Let $\alpha \in H_n$ and let $h_\alpha \in \mathscr{H}$ be defined by

$$h_\alpha(\sigma v_i) = \begin{cases} 0 & n < i \\ \sigma \pi^{i+1} \varphi^{i-n} T_{n,i}(\alpha) & 0 \leqq i \leqq n \end{cases}$$

for $\sigma \in G_\infty$. Let $g \in T \cdot I$. Then using (1), and the fact that $T_{i,i-1}(h_\alpha(v_i)) = \pi \varphi h_\alpha(v_{i-1})$ for $1 \leqq i \leqq n$, we have, $L_n g,\ \pi^{n+1}\alpha\rangle_n$

$$\langle L_n g, h_\alpha(v_n)\rangle_n = \langle \varepsilon_n g, h_\alpha(v_n)\rangle_n + \langle L_{n-1} g, h_\alpha(v_{n-1})\rangle_{n-1}.$$

Iterating this we get

$$\begin{aligned} \langle L_n g, \pi^{n+1}\alpha\rangle_n &= \sum_{i=0}^{n} \langle \varepsilon_i g, h_\alpha(v_i)\rangle_i \\ &= T_{H/K}\left(\sum_{i=0}^{n} T_i(g(v_i)\, h_\alpha(v_i))\right) \\ &= T_{H/K} \int_{\mathfrak{F}} g \cdot h_\alpha, \end{aligned}$$

by (III, 7). From (2) we deduce that $\alpha \in \mathfrak{X}_n$ if and only if

$$T_{H/K} \int_{\mathfrak{F}} g \cdot h_\alpha \equiv 0 \bmod \pi^{n+1} \mathcal{O}_K \tag{4}$$

for all $g \in A'$.

Suppose now that there exists an $f_\alpha \in T^{-1} \cdot I$ such that $\operatorname{Res}_0 f_\alpha \in \mathcal{O}_K$ and $f_\alpha(v_i) = \varphi^n h_\alpha(v_i)$ for $0 \leqq i \leqq n$. Then if $g \in A'$ we have by Proposition 7, setting $f'_\alpha = \varphi^{-n} f_\alpha$,

$$\frac{d}{dT}(g)(0) \cdot \operatorname{Res}_0 f'_\alpha = \int_{\mathfrak{F}} g \cdot f'_\alpha \equiv \int_{\mathfrak{F}} g \cdot h_\alpha \bmod \pi^{n+1} \mathcal{O}_H.$$

The last congruence follows from (III, 3) and the fact that $T_m(\mathcal{O}_m) \subseteq \pi^m \mathcal{O}_H$. Hence, h_α satisfies (4) as our hypothesis on g and f imply that the value on the left lies in $(1-\varphi)\mathcal{O}_K$. Therefore α belongs to $\mathfrak{X}_n$.

Conversely, suppose $\alpha \in \mathfrak{X}_n$, then (4) implies in particular,

$$T_{H/K}\Big(a \cdot \int_{\mathfrak{F}} g \cdot h_\alpha\Big) = T_{H/K} \int_{\mathfrak{F}} a \cdot g \cdot h_\alpha \equiv 0 \bmod \pi^{n+1} \mathcal{O}_K.$$

for all $a \in \mathcal{O}_H$ and all $g \in T^2 \cdot I$. This implies

$$\int_{\mathfrak{F}} g \cdot h_\alpha \equiv 0 \bmod \pi^{n+1} \mathcal{O}_H$$

for all $g \in T^2 \cdot I$, since H is unramified over K. Therefore $T \cdot h_\alpha \in \mathscr{H}_n$, and so by Theorem 8, there exists an $f'_\alpha \in T^{-1} \cdot I$ such that $f'_\alpha(u) = h_\alpha(u)$ for $u \in \mathfrak{F}_0$. We still need to investigate the residue of f'_α. By (4) with $g(T) = (1-\varphi) b \cdot T$, and Proposition 7, we have,

$$T_{H/K}((1-\varphi) b \cdot \operatorname{Res}_0 f'_\alpha) \equiv 0 \bmod \pi^{n+1} \mathcal{O}_K,$$

for all $b \in \mathcal{O}_H$. It is easy to see that this implies that $\operatorname{Res}_0 f'_\alpha = c + \pi^{n+1} d$, where $c \in \mathcal{O}_K$ and $d \in \mathcal{O}_H$. Let

$$f_\alpha = \varphi^n(f'_\alpha(T) - d(T^{-1}[\pi^{n+1}](T))).$$

It now follows that f_α satisfies the conditions of our theorem and the proof is complete.

Corollary 27 (Iwasawa). *With notation as in Corollary* 25: *If*

$$\alpha_n = p^{-(n+1)} \frac{\zeta_n}{\zeta_n - 1}$$

$$\beta_n = p^{-(n+1)} \sum_{i=0}^{n} \zeta_n^{p^i}$$

then α_n *and* $(1-\sigma)\beta_n$ *are elements of* $\mathfrak{X}_n$ *for all* $\sigma \in \operatorname{Gal}(\mathbb{Q}_p(\zeta_n)/\mathbb{Q}_p)$.

Proof. Let

$$g(T) = \frac{T+1}{T} \quad \text{and} \quad h(T) = \sum_{i=0}^{\infty} [p^i](T).$$

This series converges by Lemma 20(ii). Let $\zeta_i = \zeta_n^{p^{n-i}}$, and define α_i, β_i analogously to α_n and β_n. Let $\tilde{\sigma} \in G_\infty$, such that $\tilde{\sigma}/H_n = \sigma$. It is then clear that,

$$\varepsilon_i g = g(\zeta_i - 1) = p^{i+1} \alpha_i$$
$$\varepsilon_i((1-\tilde{\sigma}) h) = p^{i+1}(1-\sigma)\beta_i.$$

An easy computation shows $T_{n,i}(\alpha_n) = \alpha_i$ and $T_{n,i}(\beta_n) = \beta_i$. Hence the corollary follows from the previous theorem with α equal to either α_n or β_n and f_α equal to either g or f. (See remark at the end of this section.)

Recall that $\mathfrak{X}_\infty = \varprojlim \mathfrak{X}_n$, the inverse limit being taken with respect to the trace maps; also recall that $\mathcal{S}$ is the homomorphism introduced in III. We then have:

Theorem 28. *There exists a unique map* $\theta(=\theta_v)$ *from* $\mathfrak{X}_\infty$ *into* $\mathcal{O}_H((T))$ *such that the following diagrams commute for all* $n \geqq 0$.

$$\begin{array}{ccc} \mathfrak{X}_\infty & \xrightarrow{\theta} & \mathcal{O}_H((T)) \\ \downarrow{\scriptstyle T_{\infty,n}} & & \downarrow{\scriptstyle \varepsilon_n} \\ H_n & \xrightarrow{\pi^{n+1}\varphi^n} & H_n \end{array}$$

Proof. The proof is exactly analogous to the proof of Theorem 16, using Theorem 26 in place of Theorem 15.

Corollary 29. θ *defines a topological* $\mathcal{A}_\infty$*-isomorphism from* $\mathfrak{X}_\infty$ *onto* $\{f \in \mathcal{O}_H((T)) \colon \mathcal{S}(f) = \pi\varphi f\}$.

Proof. Let $\alpha\in\mathfrak{X}_\infty$. If $f=\theta(\alpha)$ then

$$\varepsilon_i\mathscr{S}(f)=T_{i+1,i}(\varepsilon_{i+1}(f))=\varepsilon_i(\pi\varphi(f)).$$

Hence by our uniqueness principle $\mathscr{S}(f)=\pi\varphi f$. On the other hand if $g\in\mathcal{O}_H((T))$ and $\mathscr{S}(g)=\pi\varphi g$ then since $\mathscr{S}(g)_\pi$ has the same polar part as g by (III, 2) we see easily that $g\in T^{-1}\cdot I$ and that $\operatorname{Res}_0 g\in\mathcal{O}_K$. Thus by Theorem 26 $c_g=\{\pi^{-(n+1)}\varphi^{-(n+1)}\varepsilon_n g\}\in\mathfrak{X}_\infty$; and clearly the map $g\mapsto c_g$ is the inverse of θ. The remaining statements may be proven using arguments similar to those used in the proof of Corollary 17.

Remark. This theorem relates the modules $\mathfrak{X}_n$ to an eigenspace of the operator $\varphi^{-1}\mathscr{S}$. In the height one theory alluded to after Corollary 25 we show how the other eigenspaces are related to modules with interesting properties. For instance, the image of λ, and $\operatorname{Log}(1+T)$ correspond to the eigenspaces with eigenvalues $\pi^{-1}p$, and 1, respectively. In fact if f is as in Corollars 25, and, g and h are as in Corollary 27 then,

$$\mathscr{S}(\tilde{f})=\tilde{f},\qquad \delta(g)=p\cdot g\quad\text{and}\quad \mathscr{S}((1-\tilde{\sigma})h)=p(1-\tilde{\sigma})h,$$

where $\tilde{f}=(1-a^{-1})p(1-p)^{-1}+f$. Using the above ideas we show (in the situation of Corollary 27) that

$$\mathfrak{X}_n=\mathscr{A}_\infty\alpha_n+\mathscr{I}_0\beta_n$$

where $\mathscr{I}_0$ is the closed ideal in $\mathscr{A}_\infty$ generated topologically by $\{1-\sigma\}_{\sigma\in G_\infty}$. This result was originally proven, by Iwasawa [I₂], using the explicit reciprocity law.

Remark. Let $\delta\colon\mathscr{M}\to\mathcal{O}_H((T))$ be the intrinsic logarithmic derivation,

$$\delta\colon f\mapsto\frac{1}{\lambda'}\frac{f'}{f}.$$

$\left(\text{Here } g'=\dfrac{d}{dT}g.\right)$ It is not difficult to see that

$$\pi\delta\mathscr{N}f=\mathscr{S}\delta f.$$

Therefore, if $f\in\mathscr{M}_\infty$ then $\mathscr{S}(\delta f)=\pi\varphi\delta f$. By Theorem 16 and the previous theorem, we see that this allows us to define a homomorphism ψ from X_∞ into $\mathfrak{X}_\infty$. This homomorphism is intimately connected with the class field theory of H_∞ and will be discussed more fully in a subsequent paper.

References

[C] Coates, J., Wiles, A.: On p-adic L-functions and Elliptic Units. Journal Australian Math. Soc., A, **25**, 1–25 (1978)

[F] Fontaine, J.M.: Corps de Séries Formelles et Extensions Galoisiennes des Corps Locaux, Séminaire de Théorie des Nombres de Grenoble, 28–38 (1971–72)

[Fr] Fröhlich, A.: Formal Groups, Lecture Notes in Mathematics, No. 74. Berlin: Springer-Verlag 1968

[I_1] Iwasawa, K.: Explicit Formulas for the Norm Residue Symbol. Jour. Math. Soc. Japan **20**, 151–164 (1968)

[I_2] Iwasawa, K.: On Some Modules in the Theory of Cyclotomic Fields. Jour. Math. Soc. Japan **20**, 42–82 (1964)

[L] Lubin, J., Tate, J.: Formal Complex Multiplication in Local Fields. Ann. of Math. **81**, 380–387 (1965)

[W] Wiles, A.: Higher Explicit Reciprocity Laws. Ann. of Math. **107**, 235–254 (1978)

Received April 2, 1978/March 5, 1979

Inventiones math. 53, 117–133 (1979)

Inventiones mathematicae

An Arithmetic Characterization of the Rational Homotopy Groups of Certain Spaces

John B. Friedlander and Stephen Halperin

University of Toronto, Scarborough College, Physical Science Group, West Hill, Ontario, M1C 1A4, Canada

1. Introduction

In this paper we consider path connected topological spaces S which satisfy the conditions:

(1.1) $\dim H^*_{\mathrm{sing}}(S;\mathbb{Q})<\infty$ and $\dim \pi^*_\psi(S)<\infty$;

we call such spaces *spaces of type F*. For simply connected spaces $\pi^*_\psi(S)$ coincides with $\mathrm{Hom}_{\mathbb{Z}}(\pi_*(S);\mathbb{Q})$ (see § 4) so that in this case the condition reduces to

(1.2) $\dim H^*_{\mathrm{sing}}(S;\mathbb{Q})<\infty$ and $\dim \pi_*(S)\otimes\mathbb{Q}<\infty$.

The general definition of $\pi^*_\psi(S)$, to be given in § 4, is in terms of the minimal model of S.

Spaces of type F include all homogeneous spaces G/K in which K is a closed connected subgroup of a connected Lie group G (cf. [4], [5; Chap. 11]). If $E\to B$ is a Serre fibration with path connected fibre E_b and 1-connected base B and if E_b and B are of type F, so is E (apply [8; Theorems 20.3, 10.1, 10.3 and 2.2]). Conversely if $\dim H^*_{\mathrm{sing}}(E_b;\mathbb{Q})<\infty$ and $\dim \pi^*_\psi(E)<\infty$ then [10; Theorem 4.15 iv)] asserts that E_b is of type F.

Let S be a space of type F and let $2b_1-1,\dots,2b_q-1$; $2a_1,\dots,2a_r$ be the (positive) integers occurring as the degrees of a homogeneous basis of $\pi^*_\psi(S)$. The integers $b_1,\dots,b_q$; $a_1,\dots,a_r$ will be called the *b-exponents* and the *a-exponents* of S. If S is one-connected these integers describe the locations of the rational homotopy groups.

The main purpose of this paper is to give an arithmetic characterization of the sequences of positive integers which can occur as the b- and a-exponents of a space of type F. To state this we need:

Definition. Let $B=(b_1,\dots,b_q)$, $A=(a_1,\dots,a_r)$ be two finite sequences of positive integers. We say that (B,A) satisfies S.A.C. (strong arithmetic condition) if, for

0020-9910/79/0053/0117/$03.40

every subsequence A^* of A of length s $(1 \leqq s \leqq r)$ there exist at least s elements b_j of B of the form

$$b_j = \sum_{a_i \in A^*} \gamma_{ij} a_i,$$

where the γ_{ij} are non-negative integers, and $\sum_{a_i \in A^*} \gamma_{ij} \geqq 2$.

If the requirement that $\sum_{a_i \in A^*} \gamma_{ij} \geqq 2$ is dropped, we say that (B, A) satisfies A.C.

Note that either condition ensures that $q \geqq r$. (Choose $A^* = A$.) We can now state the first theorem.

Theorem 1. *Let $B = (b_1, \ldots, b_q)$ and $A = (a_1, \ldots, a_r)$ be a pair of sequences of positive intgers. The following conditions are equivalent:*

1) *(B, A) satisfies S.A.C.*
2) *The sequences B and A occur as the b- and a-exponents of a space S of type F.*

Moreover, if $b_j \geqq 2$ for all j and S.A.C. holds, then S may be chosen to be simply connected; if in addition $q > r$, S may be taken to be a closed manifold.

1.3. **Corollary.** *Let S be a one-connected space satisfying* (1.2). *Let $n(S)$ be the largest integer m such that $H^m_{\mathrm{sing}}(S; \mathbb{Q}) \neq 0$. Then,*

(1) $\dim \pi_*(S) \otimes \mathbb{Q} \leqq n(S)$.

(2) $\sum_k 2k \dim(\pi_{2k}(S) \otimes \mathbb{Q}) \leqq n(S)$.

In particular, $\pi_{2k}(S)$ is a torsion group for $2k > n(S)$.

(3) $\sum_k (2k-1) \dim(\pi_{2k-1}(S) \otimes \mathbb{Q}) \leqq 2n(S) - 1$.

In particular, $\pi_{2k-1}(S)$ is a torsion group for $k > n(S)$.

1.4. *Remark.* The example S^{2k} shows that (2) and (3) are sharp.

Our second theorem is an application to transformation groups. Recall that a compact connected Lie group G of rank k has b-exponents $c_1, \ldots, c_k$ and no a-exponents. (This follows from the fact that $H^*(G; \mathbb{Q})$ is an exterior algebra on generators of degree $2c_i - 1$ $(1 \leqq i \leqq k)$; see the definition in § 4.)

Theorem 2. *Let S be a Hausdorff space of type F with exponents $b_1, \ldots, b_q$; $a_1, \ldots, a_r$ and assume that S is either compact, or paracompact with $\mathrm{cd}_{\mathbb{Q}}[S] < \infty$* (cf. [13]). *Assume a compact connected Lie group G with b-exponents $c_1, \ldots, c_k$ acts continuously on S so that all the isotropy subgroups are finite. Then, the pair of sequences*

$$B = (b_1, \ldots, b_q), \qquad A' = (a_1, \ldots, a_r, c_1, \ldots, c_k)$$

satisfies A.C.

1.5. *Remark.* Theorem 2 asserts in particular that $k \leqq q-r$; this is precisely Theorem 1 of [1].

1.6. *Examples.* 1. The pair of sequences $B=(2,2,2,7,8)$; $A=(4)$, satisfies S.A.C. and is thus the exponents of a closed manifold M. On the other hand the group $SU(n)$ has exponents $(2,3,4,5,\ldots,n)$ and hence, if $n \geqq 3$, cannot act on any such M with finite isotropy subgroups. A similar calculation shows that $SO(n)$, $n \geqq 5$, and $Sp(n)$, $n \geqq 2$, cannot act on M with finite isotropy subgroups.

2. If $SU(n)$ acts on a finite product of odd spheres, all of which have dimension $<2n-1$, then some isotropy subgroup has positive dimension.

Although the two theorems stated above are topological in nature, they are deduced from a theorem in algebraic geometry (Theorem 3) whose proof constitutes the major part of the paper. This arises in the following way.

By using the theory of minimal models we show that a pair of sequences (B, A) give the exponents of a space of type F if and only if there are homogeneous polynomials (with no linear terms) $f_1, \ldots, f_q$ of degrees $b_1, \ldots, b_q$ in variables $u_1, \ldots, u_r$ of degrees $a_1, \ldots, a_r$ such that the f_i generate an ideal of finite codimension in $\mathbb{Q}[u_1, \ldots, u_r]$. This raises the question of when sets Φ_i of monomials can be combined (linearly) into polynomials f_i with this property, and the answer is the content of Theorem 3, which we give next.

Consider $R=k[u_1, \ldots, u_r]$, the ring of polynomials in r variables u_i over an *infinite* field k. Assume

$$\Phi_i = \{\sigma_{ij}\}_{j=1,\ldots,l_i}, \qquad i=1,\ldots,q$$

are families of monomials σ_{ij} in the variables $u_1, \ldots, u_r$.

Definition. We say the families $\Phi_1, \ldots, \Phi_q$ satisfy P.C. (polynomial condition) if for each s and for each set of s variables $u_{i_1}, \ldots, u_{i_s}$ there are at least s families $\Phi_{m_1}, \ldots, \Phi_{m_s}$ containing a monomial in $k[u_{i_1}, \ldots, u_{i_s}]$.

Remark. Clearly if $\Phi_1, \ldots, \Phi_q$ satisfy P.C. then $q \geqq r$.

Given sets $\Phi_1, \ldots, \Phi_q$ of monomials, we consider polynomials $f_1, \ldots, f_q \in k[u_1, \ldots, u_r]$ of the form

$$(1.7) \quad f_i = \sum_{j=1}^{l_i} c_{ij}\sigma_{ij}, \qquad c_{ij} \in k, \ \sigma_{ij} \in \Phi_i \ (1 \leqq j \leqq l_i)$$

Theorem 3. *Assume that $\Phi_1, \ldots, \Phi_q$ are sets of monomials as above. Then $\Phi_1, \ldots, \Phi_q$ satisfy P.C. if and only if there are polynomials $f_1, \ldots, f_q$ of the form* (1.7) *such that*

$$(1.8) \quad \dim_k k[u_1, \ldots, u_r]/(f_1, \ldots, f_q) < \infty.$$

1.9. *Remark.* A geometric formulation of Theorem 3 is given by noting that (1.8) holds if and only if the variety $V((f_1, \ldots, f_q))$ has at most finitely many points in $\bar{k}^r$, $\bar{k}$ being the algebraic closure of k.

The connection between P.C. and the earlier conditions occurs as follows.

Let $B=(b_1, \ldots, b_q)$, $A=(a_1, \ldots, a_r)$ be finite sequences of positive integers. To the variable $u_i (1 \leqq i \leqq r)$ assign the degree a_i. For $1 \leqq j \leqq q$, denote by Φ_j (respectively Φ'_j) the set of monomials (resp. non-linear monomials) in $u_1, \dots, u_r$ of degree b_j. Observe that $\Phi_1, \dots, \Phi_q$ (resp. $\Phi'_1, \dots, \Phi'_q$) satisfies P.C. if and only if (B, A) satisfies A.C. (resp. S.A.C.). There follows at once from Theorem 3:

1.10. **Corollary.** *In $k[u_1, \dots, u_r]$, there exist $f_1, \dots, f_q$; f_j a linear combination of monomials (resp. non-linear monomials) of degree b_j, with*

$$\dim k[u_1, \dots, u_r]/(f_1, \dots, f_q) < \infty,$$

if and only if (B, A) satisfies A.C. (resp. S.A.C.).

The paper is organized as follows. § 2 and § 3 contain some elementary consequences of Theorem 1 (including the proof of Corollary 1.3). In § 4 we recall some facts from the theory of minimal models and use them to deduce Theorems 1 and 2 from Corollary 1.10. § 5–§ 7 are devoted to the proof of Theorem 3 (a somewhat stronger form is given for the case $q = r$). § 8 contains the proof of Proposition 2.8 which gives, for our spaces, an upper bound for the Betti numbers in terms of the b- and a-exponents.

We remark, finally, that although, in the topological sections, the coefficient field is taken to be the rational field $\mathbb{Q}$, the results apply verbatim to any field of characterstic zero.

2. Some Topological Consequences

Spaces of type F satisfy very rigid restrictions (in addition to those imposed by Theorem 1). For example, if S has type F, then [9; Theorem 3] $H^*_{\text{sing}}(S; \mathbb{Q})$ is a Poincaré duality algebra. Thus, if $n(S)$ (called the *formal dimension* of S) is the largest integer m such that $H^m_{\text{sing}}(S; \mathbb{Q}) \neq 0$, then

$$\dim H^p_{\text{sing}}(S; \mathbb{Q}) = \dim H^{n(S)-p}_{\text{sing}}(S; \mathbb{Q}), \qquad p = 0, 1, \dots.$$

Moreover, if $b_1, \dots, b_q$ and $a_1, \dots, a_r$ are the b- and a-exponents of S then [9; Theorem 1] shows that

$$(2.1) \quad q = r \Leftrightarrow \chi(S) \neq 0 \Leftrightarrow \dim H^p_{\text{sing}}(S; \mathbb{Q}) = 0, \qquad p \text{ odd}.$$

Furthermore, [9; Theorem 3] gives

$$(2.2) \quad n(S) = 2\left(\sum_1^q b_i - \sum_1^r a_i\right) - (q - r).$$

Finally, [9; Corollary 2 to Theorem 5] shows that if $q = r$ then

$$(2.3) \quad \frac{\prod_{i=1}^{r} (1 - t^{2b_i})}{\prod_{i=1}^{r} (1 - t^{2a_i})} = \sum_0^{n(S)} \dim H^p_{\text{sing}}(S; \mathbb{Q})\, t^p,$$

and

$$(2.4)\quad \frac{\prod_{i=1}^{r} b_i}{\prod_{i=1}^{r} a_i} = \dim H^*_{\mathrm{sing}}(S;\mathbb{Q}).$$

Formulae of type (2.3) and (2.4) originate in the work of Cartan [4] and Koszul [11].

We now use Theorem 1 to obtain further properties of spaces of type F.

2.5. **Lemma.** *Let* $B=(b_1,\ldots,b_q)$; $b_1 \geqq b_2 \geqq \ldots \geqq b_q$; *and* $A=(a_1,\ldots,a_r)$; $a_1 \geqq a_2 \geqq \ldots \geqq a_r$. *If* (B,A) *satisfies S.A.C. then* $b_i \geqq 2a_i$ *for* $1 \leqq i \leqq r$.

Proof. Fix i with $1 \leqq i \leqq r$ and consider integers of the form $\sum_{j=1}^{i} \gamma_j a_j$ with non-negative integers γ_j satisfying $\sum_j \gamma_j \geqq 2$. It is clear that any such integer is $\geqq 2a_i$. On the other hand, by S.A.C. integers of this form include at least i of the b's, at least one of which must necessarily be $\leqq b_i$. q.e.d.

Lemma 2.5, when combined with (2.2) and (2.4) yields:

2.6. **Proposition.** Let S be a space of type F with exponents $b_1,\ldots,b_q$; $a_1,\ldots,a_r$. Then,

(1) $n(S) \geqq 2\sum_1^r a_i + (q-r) \geqq q+r$.

(2) $n(S) \geqq \sum_1^q b_i$.

(3) If $q=r$, then $\dim H^*_{\mathrm{sing}}(S;\mathbb{Q}) \geqq 2^r$.

2.7. *Remarks.* 1. Corollary 1.3 follows from 2.6.

2. The statements of 2.5 and 2.6 are sharp in the case $q=r$, $b_i = 2a_i$, $1 \leqq i \leqq r$.

In the case $q=r$, (2.3) gives an expression for the Betti numbers in terms of the b and a-exponents. For the general case we have the following weaker result (to be proved in § 8).

2.8. **Proposition.** Let S be a space as in Proposition 2.6.

(1) The expression $\prod_1^q (1-t^{2b_i})/(1-t)^{q-r}\prod_1^r (1-t^{2a_i})$ is a polynomial $\sum_0^{n(S)} c_m t^m$ with non-negative integral coefficients.

(2) For $0 \leqq m \leqq n(S)$, we have

$\dim H^m_{\mathrm{sing}}(S;\mathbb{Q}) \leqq c_m$.

2.9. **Corollary.** $\dim H^*_{\mathrm{sing}}(S;\mathbb{Q}) \leqq 2^{q-r}\dfrac{\prod b_i}{\prod a_j} \leqq (2n(S))^{n(S)}$.

Remark. (1) of (2.8) is due to C. Allday (private communication). We give instead a short proof based on Theorem 1 but make full use of Allday's technique (cf. Proposition 8.1) in our proof of (2).

3. An Arithmetic Consequence

3.1. **Proposition.** If $B=(b_1,\dots,b_q)$, $A=(a_1,\dots,a_r)$ and (B,A) satisfies A.C., then

$$f(t)=\prod_{j=1}^{q}(1+t+\dots+t^{b_j-1})\Big/\prod_{i=1}^{r}(1+t+\dots+t^{a_i-1})$$

is a polynomial with non-negative integral coefficients.

Proof. Let $A'=(a_1,\dots,a_q)$ where $a_{r+1}=\dots=a_q=1$. Since (B,A') satisfies A.C., it suffices to consider the case $q=r$.

An elementary argument shows that if we cancel a pair for which $a_i=b_j$, A.C. is not destroyed. By cancelling such pairs, we reduce to the case that (B,A) satisfy S.A.C. Choose a space S of type F with $b_1,\dots,b_q$; $a_1,\dots,a_q$ as exponents.

By (2.3), we have

$$f(t^2)=\sum_{0}^{n(S)}\dim H^p_{\mathrm{sing}}(S;\mathbb{Q})\,t^p. \quad \text{q.e.d.}$$

Remarks. 1. For $q=r=1,2$, the converse of 3.1 is also true (cf. [6], [14]). If $r>2$ this is no longer so, as shown by the example $B=(6,7,20)$, $A=(3,4,10)$. Notice that, in this example, all the conclusions of 2.5 and 2.6 fail! (Here there is no space S, and $n(S)$ must be replaced by $2\deg f$.)

2. If in the definition of A.C., the integers γ_{ij} are permitted to be negative, the modified condition is easily seen to be equivalent to the requirement that f be a polynomial.

3. If we demand only that f be a polynomial, we can still get a bound of the form $\phi(b_0)\leqq\deg f$ where $b_0=\max\limits_{1\leqq i\leqq r} b_i$ and ϕ is the Euler totient. For the cyclotomic polynomials this bound is sharp. Thus, since $b/\phi(b)$ is unbounded, we can have no bounds of the form 1) or 2) of 2.6, even allowing an arbitrary constant factor.

4. Minimal Models

A commutative graded differential algebra (c.g.d.a.) over $\mathbb{Q}$ is a graded differential algebra (A,d) with $A=\sum_{p\geqq 0}A^p$ and $ab=(-1)^{pq}ba$, $a\in A^p$, $b\in A^q$. Its cohomology is denoted by $H(A)$. A is called *connected* if $A^0=\mathbb{Q}$.

If $X=\sum_{p\geqq 0}X^p$ is a graded space then ΛX denotes the tensor product:

$$\Lambda X=\text{Exterior algebra }(X^{\text{odd}})\otimes\text{Symmetric algebra }(X^{\text{even}}).$$

The ideal generated by X is denoted by Λ^+X.

A *KS (Koszul-Sullivan) complex* is a c.g.d.a. of the form $(\Lambda X, d)$ with the property that for some well ordered homogeneous basis $\{x_\alpha\}_{\alpha\in\mathscr{I}}$ of X:

(4.1) $dx_\alpha \in \Lambda X_{<\alpha}, \qquad \alpha\in\mathscr{I}.$.

(Here $\Lambda X_{<\alpha}$ is the subalgebra generated by the x_β, $\beta<\alpha$.) A KS complex is called minimal if

$$\beta<\alpha \Rightarrow \deg x_\beta \leqq \deg x_\alpha, \qquad \beta,\alpha\in\mathscr{I}.$$

If ΛX is connected this is equivalent to

$$\operatorname{Im} d \subset \Lambda^+ X\cdot \Lambda^+ X.$$

In [15; § 7] Sullivan defines a contravariant functor from topological spaces to c.g.d.a.'s, denoted by $S \rightsquigarrow (A(S), d)$, such that there are natural multiplicative isomorphisms, $H^*_{\text{sing}}(S;\mathbb{Q})\cong H(A(S),d)$ (cf. also [7; Chap. 15], [12; Chap. 3], [3]).

He then shows ([15; § 5], [12; Théorème II.6], [3], [7; Chap. 6]) that if S is path connected there is a connected minimal KS complex $(\Lambda X, d)$, uniquely determined up to isomorphism, and a homomorphism

$$m_s\colon (\Lambda X, d)\to (A(S), d)$$

such that $m_s^*\colon H(\Lambda X)\to H(A(S))$ is an isomorphism. The KS complex $(\Lambda X, d)$ is called the *minimal model of S*.

If $(\Lambda X, d)$ is the minimal model of S, we put

$$\pi^*_\psi(S)=\Lambda^+ X/\Lambda^+ X\cdot \Lambda^+ X;$$

it is the graded space of indecomposable elements. Notice that the inclusion $X\to\Lambda^+ X$ gives an isomorphism $X\cong\pi^*_\psi(S)$.

The results cited above provide a multiplicative isomorphism

$$H(\Lambda X)\cong H^*_{\text{sing}}(S;\mathbb{Q}).$$

By another result of Sullivan ([15; Theorem 10.1], [12; Théorème IV.8], [3]), if S is nilpotent (in particular, if S is 1-connected) and

$$\dim H^p_{\text{sing}}(S;\mathbb{Q})<\infty, \qquad p=1,2,\ldots, \text{ then}$$

(4.2) $X^p\cong \pi^p_\psi(S)\cong \operatorname{Hom}(\pi_p(S);\mathbb{Q}), \qquad p\geqq 2.$

Conversely ([15; Theorem 10.2], [12; Théorème IV.8], [3]), if $(\Lambda X, d)$ is any connected minimal KS complex such that $\dim X^p<\infty$, $p=1,2,\ldots,$ then it is the minimal model of a path connected CW complex. If $X^1=0$, the CW complex is 1-connected.

With these observations we can prove Theorems 1 and 2, with the aid of Corollary 1.10.

Proof of Theorem 1. Let $B=(b_1,\dots,b_q)$, $A=(a_1,\dots,a_r)$ be sequences of positive integers. If they are the exponents of a space S of type F, then the minimal model $(\Lambda X,d)$ of S satisfies

(4.3) $\dim H(\Lambda X,d)<\infty$

and

(4.4) X has a finite homogeneous basis $x_1,\dots,x_q$, $y_1,\dots,y_r$ with $\deg x_i=2b_i-1$, $\deg y_j=2a_j$.

Conversely, let $(\Lambda X,d)$ be a minimal KS complex satisfying (4.3) and (4.4). By the remarks above there is a CW complex S with $(\Lambda X,d)$ as minimal model. Now (4.3) and (4.4) imply that S has type F and that its b and a-exponents are B and A. Moreover, if $b_i\geqq 2$, $1\leqq i\leqq q$, then $X^1=0$ and so S may be chosen simply connected. If in addition $q>r$, then [9; Theorems 3 and 4] imply that $H^*(S)$ is a Poincaré duality algebra with hyperbolic Poincaré scalar product. Thus a theorem of Sullivan ([15; Theorem 13.2], [2; Théorème I, Partie II] asserts that S may be chosen to be a simply connected closed manifold.

We have now reduced Theorem 1 to the assertion that (B,A) satisfies S.A.C. if and only if there is a minimal KS complex $(\Lambda X,d)$ satisfying (4.3) and (4.4).

Assume first that (B,A) satisfies S.A.C. In $\mathbb{Q}[y_1,\dots,y_r]$ assign y_i the degree $2a_i$. By Corollary 1.10 there are polynomials $f_1,\dots,f_q\in\mathbb{Q}[y_1,\dots,y_r]$, having no linear term, such that f_j is homogeneous of degree $2b_j$ and

$$\dim_{\mathbb{Q}}\mathbb{Q}[y_1,\dots,y_r]/(f_1,\dots,f_q)<\infty.$$

Let X be the graded space with basis $x_1,\dots,x_q$, $y_1,\dots,y_r$ and $\deg x_i=2b_i-1$. Define a minimal KS complex $(\Lambda X,d)$ by setting $dy_i=0$, $dx_i=f_i$. Then it is a trivial fact (cf. [9; end of § 3]) that $\dim H(\Lambda X)<\infty$. Thus $(\Lambda X,d)$ satisfies (4.3) and (4.4).

Conversely, suppose $(\Lambda X,d)$ is a minimal KS complex satisfying (4.3) and (4.4). (Note that $\Lambda(y_1,\dots,y_r)=\mathbb{Q}[y_1,\dots,y_r]$.) Accoring to [9; Proposition 1] there is a second differential $\bar d$ in ΛX, with the following properties:

(4.5) $H(\Lambda X,\bar d)<\infty$

(4.6) $\bar d y_i=0,\quad 1\leqq i\leqq r$

(4.7) $\bar d x_j\in\Lambda^+(y_1,\dots,y_r)\cdot\Lambda^+(y_1,\dots,y_r),\quad 1\leqq j\leqq q.$

Clearly the inclusion $\Lambda(y_1,\dots,y_r)\to\Lambda X$ induces an inclusion

$$\Lambda(y_1,\dots,y_r)/(\bar d x_1,\dots,\bar d x_q)\to H(\Lambda X,\bar d).$$

Thus $\dim\Lambda(y_1,\dots,y_r)/(\bar d x_1,\dots,\bar d x_q)<\infty$ and Corollary 1.10 implies that (B,A) satisfy S.A.C. q.e.d.

Proof of Theorem 2. Let $(\Lambda X,d)$ be the minimal model for S. It is shown in [1; § 3] that under the hypotheses of Theorem 2 there is a KS complex $(\Lambda Q_G\otimes\Lambda X,D)$ such that

(4.8) $\dim H(\Lambda Q_G \otimes \Lambda X, D) < \infty$.

(4.9) Q_G has a homogeneous basis $z_1, \ldots, z_k$ with $\deg z_i = 2c_i$.

(4.10) $D z_i = 0$ and $D(1 \otimes \Phi) - 1 \otimes d\Phi \in \Lambda^+ Q_G \otimes \Lambda X, \quad \Phi \in \Lambda X$.

Although $(\Lambda Q_G \otimes \Lambda X, D)$ may not be minimal, because of (4.10) a trivial modification of the argument of [9; Proposition 1] yields a second differential $\bar{D}$ in $\Lambda Q_G \otimes \Lambda X$ such that

$$H(\Lambda Q_G \otimes \Lambda X, \bar{D}) < \infty$$

and

$$\bar{D} z_i = \bar{D} y_j = 0, \qquad 1 \leqq i \leqq k, \; 1 \leqq j \leqq r$$
$$\bar{D} x_l \in \Lambda Q_G \otimes \Lambda(y_1, \ldots, y_r), \qquad 1 \leqq l \leqq q.$$

As in the proof of Theorem 1 we can now apply Corollary 1.10 to obtain that $(b_1, \ldots, b_q;\, a_1, \ldots, a_r,\, c_1, \ldots, c_k)$ satisfy A.C. q.e.d.

5. Noetherian Rings

Let R be an integral domain, finitely generated as an algebra over an infinite field k. Every proper ideal $I \subset R$ is uniquely the irredundant intersection of finitely many primary ideals $\mathcal{Q}_j$ [16, Theorem 4, p. 209]; the prime ideals P_j associated with these primary ideals are called the *associated primes* of I and P_j is called *isolated* if it does not properly contain any other associated P_i.

The transcendence degree of R/P_j is called the *dimension of* P_j (Dim P_j) and the maximum of the dimensions of the isolated primes is called the *dimension of* I (Dim I).

5.1. **Lemma.** *If r is the transcendence degree of R and I is generated by s elements then each isolated prime P_j has dimension $\geqq r-s$. In particular.*

$$\operatorname{Dim} I \geqq r - s.$$

Moreover, $\operatorname{Dim} I = r - s \Leftrightarrow \operatorname{Dim} P_j = r - s$ *for each isolated prime P_j of I.*

Proof. [17; Theorem 22, p. 196.]

5.2. **Lemma.** *For I, a proper ideal of R,*

$$\dim R/I < \infty \Leftrightarrow \operatorname{Dim} P_j = 0 \quad \textit{for all associated isolated primes}$$
$$\Leftrightarrow \operatorname{Dim} I = 0.$$

Proof. That the last two are equivalent is immediate.

If $\operatorname{Dim} P_j = 0$, then every non-zero element of R/P_j satisfies an algebraic equation over k, and since R/P_j is an integral domain, we may take the constant term to be -1. It follows that every non-zero element is invertible whence R/P_j is an algebraic extension of k (finitely generated since R is) and $\dim R/P_j < \infty$.

If $\mathcal{Q}_j$ is the corresponding primary ideal, then $\dim R/\mathcal{Q}_j < \infty$. Supposing now that $\operatorname{Dim} P = 0$ for all isolated primes of I, then $\operatorname{Dim} P = 0$ for any associated prime and so if $\mathcal{Q}_j$ are the primary ideals for I, $\dim R/\mathcal{Q}_j < \infty$, $\forall j$. Since $I = \bigcap_j \mathcal{Q}_j$ we obtain $\dim R/I < \infty$.

Conversely if $\dim R/I < \infty$, then $\dim R/P_j < \infty$, $\forall$ associated primes P_j. Thus R/P_j is a finite dimensional integral domain over k and hence an algebraic extension, whence $\operatorname{Dim} P_j = 0$, for all j. q.e.d.

6. The Main Algebraic Result

In this section we state Theorem 6.1 which is, for the special case $q = r$, a strengthening of Theorem 3. We then prove the easier half of 6.1 and show that Theorem 3 (even for $q > r$) follows from 6.1.

Let $R = k[u_1, \ldots, u_r]$, k an infinite field, and let $\Phi_1, \ldots, \Phi_r$ be families of monomials in $u_1, \ldots, u_r$ as in § 1.

6.1. **Theorem.** (*A*) *Assume* $\Phi_1, \ldots, \Phi_r$ *satisfy P.C. There are polynomials* $f_1, \ldots, f_r \in k[u_1, \ldots, u_r]$ *of the form*

(6.2) $f_i = \sum_{j=1}^{l_i} c_{ij}\sigma_{ij}$, $\sigma_{ij} \in \Phi_i$, $c_{ij} \in k$, *such that for all*

$$1 \leqq i_1 < \ldots < i_s \leqq r,$$

$$\operatorname{Dim}(f_{i_1}, \ldots, f_{i_s}) = r - s.$$

In particular,

(6.3) $\dim_k k[u_1, \ldots, u_r]/(f_1, \ldots, f_r) < \infty$.

(*B*) *Conversely, suppose* $f_1, \ldots, f_r$ *are polynomials satisfying* (6.2) *and* (6.3). *Then* $\Phi_1, \ldots, \Phi_r$ *satisfy P.C.*

6.4. *Remark.* By Macaulay's theorem [17; Theorem 26, p. 203], each of the ideals $(f_{i_1}, \ldots, f_{i_s})$ is unmixed, i.e. each associated prime is isolated and of dimension $r - s$.

Proof of Theorem 6.1 (*B*). If P.C. fails we can, by renumbering, assume that each monomial in Φ_i, $(s \leqq i \leqq r)$ is in the ideal $(u_{s+1}, \ldots, u_r)$.

For any choice of the f_j, we have $f_j \in (u_{s+1}, \ldots, u_r)$, $j \geqq s$, so

$$(f_1, \ldots, f_r) \subset (f_1, \ldots, f_{s-1}, u_{s+1}, \ldots, u_r)$$

By Lemma 5.1, $\operatorname{Dim}(f_1, \ldots, f_{s-1}, u_{s+1}, \ldots, u_r) \geqq 1$, so, by Lemma 5.2,

$$\dim k[u_1, \ldots, u_r]/(f_1, \ldots, f_r)$$
$$\geqq \dim k[u_1, \ldots, u_r]/(f_1, \ldots, f_{s-1}, u_{s+1}, \ldots, u_r) = \infty. \quad \text{q.e.d.}$$

Proof that (6.1) *implies Theorem* 3. If (1.8) holds $\Phi_1, \ldots, \Phi_q$ satisfies P.C. by the argument above. Assume $\Phi_1, \ldots, \Phi_q$ satisfies P.C. Then, in $k[u_1, \ldots, u_q]$, so does

$\Psi_1, \ldots, \Psi_q$, where $\Psi_i = \Phi_i \cup \{u_{r+1}, \ldots, u_q\}$, $1 \leqq i \leqq q$. By Theorem 6.1, we can find polynomials $F_1, \ldots, F_q$, with F_i a linear combination of the $\sigma_{ij} \in \Phi_i$ and $u_{r+1}, \ldots, u_q$, such that $\dim k[u_1, \ldots, u_q]/(F_1, \ldots, F_q) < \infty$.

Let f_i be the polynomial obtained from F_i by putting $u_{r+1} = \ldots = u_q = 0$. q.e.d.

7. Proof of the Algebraic Theorem

We prove Theorem 6.1 (A) by an inductive argument consisting of two propositions, the former a combinatorial result giving the start of the induction, the latter giving the inductive step.

7.1. **Proposition.** *Let $U = \{u_1, \ldots, u_r\}$. Let $\Phi_1, \Phi_2, \ldots \Phi_q$ be a collection of subsets of U with the property that each subset of s of the u_j has non-void intersection with at least s of the Φ_i. Then the Φ_j can be renumbered so that $u_j \in \Phi_j$, $1 \leqq j \leqq r$.*

Remark. If, in Proposition 7.1, one interchanges the notion of set and element the resulting statement is a theorem of P. Hall [7; Theorem 1]. The two statements are clearly equivalent, and the proof we give below, while shorter, is similar to that in [7]. We thank F. Markel for drawing our attention to Hall's theorem.

Proof. Application of the hypothesis with $s = r$ shows that $q \geqq r$.

By the hypothesis with $s = 1$ we can arrange that $u_1 \in \Phi_1$. Assume (after renumbering) that $u_i \in \Phi_i$, $1 \leqq i \leqq t$. We show that (if $t < r$) we can renumber again so that $u_i \in \Phi_i$, $1 \leqq i \leqq t+1$.

Assign to each Φ_j a type number as follows:

i) Φ_j has type 1 if $j \geqq t+1$.

ii) Φ_j has type k if it does not have smaller type and if u_j is in some Φ_l of type $k-1$.

iii) Φ_j has type ∞ if it does not have finite type.

If $1 \leqq i \leqq t$ and u_i is in a set of type k, then Φ_i has type $\leqq k+1$. Thus, if Φ_i has infinite type, u_i can be in no set of finite type. If $\Phi_{i_1}, \ldots, \Phi_{i_s}$ are the sets of infinite type we apply the hypothesis to the elements $u_{i_1}, \ldots, u_{i_s}, u_{t+1}$. Now $u_{i_1}, \ldots, u_{i_s}$ can only belong to $\Phi_{i_1}, \ldots, \Phi_{i_s}$; thus u_{t+1} must be in some set Φ_{j_1} of finite type k. Relabel Φ_{j_1} as Φ_{t+1}. By definition u_{j_1} is in some set Φ_{j_2} of type $k-1$. Relabel Φ_{j_2} as Φ_{j_1}. Continue the process to get the desired relabelling. q.e.d.

Before stating Proposition 7.2 we need some notation. Write $V_i = k^{l_i}$ and define linear maps $\phi\colon V_i \to R$ by

$$\phi(c_1, \ldots, c_{l_i}) = c_1 \sigma_{i1} + \ldots + c_{l_i} \sigma_{i l_i}.$$

(Thus $\phi(V_i)$ consists of all the possible polynomials f_i of the form (6.2).)

Next, suppose we fix i and construct a new family $\Phi_i' = \{\tau_{ij}\}$, $(j = 1, \ldots, l_i)$ such that each τ_{ij} is a linear monomial (i.e. a variable) dividing σ_{ij}. If we do this for i

$=m, m+1, \ldots, r$ the resulting sequence

$$\Phi_1, \ldots, \Phi_{m-1}, \Phi'_m, \ldots, \Phi'_r$$

will be called an *m-linearization* of $\Phi_1, \ldots, \Phi_r$. Of course each such linearization determines linear maps $\phi': V_i \to R$ defined as was ϕ (and clearly $\phi(v_i) = \phi'(v_i), i<m$).

Definition. A sequence $v_1, \ldots, v_r$ $(v_i \in V_i)$ is called *m-admissible* if for all *m*-linearizations $\Phi_1, \ldots, \Phi_{m-1}, \Phi'_m, \ldots, \Phi'_r$ and for all indices $1 \leqq i_1 < \ldots < i_s \leqq r$ $(1 \leqq s \leqq r)$

$$\operatorname{Dim}(\phi'(v_{i_1}), \ldots, \phi'(v_{i_s})) = r-s.$$

7.2. **Proposition.** *Suppose $v_1, \ldots, v_r$ is m-admissible. Then for some $w \in V_m$, $v_1, \ldots, v_{m-1}, w, v_{m+1}, \ldots, v_r$ is $m+1$ admissible.*

7.3. **Lemma.** *For any $(m+1)$-linearization (Φ', ϕ') and any set of s indices $i_1 < \ldots < i_s$ not including m $(1 \leqq s \leqq r-1)$ and any isolated prime P_α of $(\phi'(v_{i_1}), \ldots, \phi'(v_{i_s}))$:*

$$\phi(V_m) \not\subset P_\alpha.$$

Proof. Note that $\phi(V_m)$ is the span of the monomials in Φ_m. We assume the lemma false and deduce a contradiction. Thus we fix an $(m+1)$-linearization (Φ', ϕ') and a set of indices $i_1 < \ldots < i_s$ (excluding m) and an isolated prime P_α of $(\phi' v_{i_1}, \ldots, \phi' v_{i_s})$ such that

$$\phi(V_m) \subset P_\alpha.$$

Then each monomial $\sigma_{mj} \in P_\alpha$ and hence has a linear factor $\tau_{mj} \in P_\alpha$. Set

$$\Phi''_m = \{\tau_{mj}\}, \qquad \Phi''_i = \Phi'_i, \qquad i>m.$$

This defines an m-linearization for which

$$\phi''(v_i) = \phi'(v_i) \qquad i \neq m$$

and

$$\phi''(v_m) \in P_\alpha.$$

Hence $P_\alpha \supset (\phi'' v_m, \phi'' v_{i_1}, \ldots, \phi'' v_{i_s})$ is a prime ideal and so P_α must contain an isolated prime $\tilde{P}$ of the ideal $(\phi'' v_m, \phi'' v_{i_1}, \ldots, \phi'' v_{i_s})$. Since $v_1, \ldots, v_r$ is m-admissible, Lemma 5.1 implies that

$$\operatorname{Dim} P_\alpha \leqq \operatorname{Dim} \tilde{P} = r-s-1.$$

On the other hand P_α is an isolated prime for $(\phi' v_{i_1}, \ldots, \phi' v_{i_s}) = (\phi'' v_{i_1}, \ldots, \phi'' v_{i_s})$ and so

$$\operatorname{Dim} P_\alpha = r-s.$$

This contradiction establishes the lemma. q.e.d.

Proof of Proposition 7.2. In Lemma 7.3 the number of linearizations, the number of sets of indices and the number of isolated primes are all finite. It follows (because k is infinite!) that

$$\phi(V_m) \not\subset \bigcup_\alpha P_\alpha,$$

where the union is over all the prime ideals arising as described in Lemma 7.3.

Choose $w \in V_m$ so that $\phi w \notin \bigcup_\alpha P_\alpha$, and let (Φ', ϕ') be any $(m+1)$-linearization.

Write

$$\bar{v}_i = \begin{cases} v_i & i \neq m \\ w & i = m. \end{cases}$$

We must show that for any $i_1 < \ldots < i_s$, every isolated prime of $(\phi' \bar{v}_{i_1}, \ldots, \phi' \bar{v}_{i_s})$ has dimension $r-s$.

Case I. m is not one of the indices $i_1, \ldots, i_s$. Choose any m-linearization (Φ'', ϕ'') extending (Φ', ϕ'); then

$$(\phi' \bar{v}_{i_1}, \ldots, \phi' \bar{v}_{i_s}) = (\phi'' v_{i_1}, \ldots, \phi'' v_{i_s})$$

and the result is true because $v_1, \ldots, v_r$ is m-admissible.

Case II. The indices can be labelled $m, j_1, \ldots, j_{s-1}$. Write

$$I = (\phi' \bar{v}_m, \phi' \bar{v}_{j_1}, \ldots, \phi' \bar{v}_{j_{s-1}})$$
$$J = (\phi' \bar{v}_{j_1}, \ldots, \phi' \bar{v}_{j_{s-1}}).$$

Choose any isolated prime P of I and an isolated prime P' of J with $P \supset P'$. Now as we observed above there is an m-linearization (Φ'', ϕ'') such that $J = (\phi'' v_{j_1}, \ldots, \phi'' v_{j_{s-1}})$ and hence

$$\operatorname{Dim} P' = r - s + 1.$$

Moreover by construction $\phi' \bar{v}_m = \phi w \notin P'$. Since $\phi' \bar{v}_m \in I \subset P$ we have $P' \subsetneqq P$ whence ([17; (2), p. 193])

$$\operatorname{Dim} P < \operatorname{Dim} P'$$

i.e. $\operatorname{Dim} P \leqq r - s$. On the other hand Lemma 5.1 gives $\operatorname{Dim} P \geqq r - s$. Thus all isolated primes of I have dimension $r-s$ and so $\operatorname{Dim} I = r - s$. q.e.d.

Proof of Theorem 6.1(*A*). For each 1-linearization $\Phi'_1, \ldots, \Phi'_r$ the set of elements $v_1, \ldots, v_r$ such that $\phi' v_1, \ldots, \phi' v_r$ are linearly independent (and hence a basis for the span of the linear monomials) is Zariski open in $V_1 \oplus \ldots \oplus V_r$.

Moreover clearly $\Phi'_1, \ldots, \Phi'_r$ satisfies P.C. and so Proposition 7.1 implies that this set is non-void. Intersecting over all possible 1-linearizations we obtain a finite intersection of non-void Zariski open sets which is then non-void because k is infinite.

If $v_1 \oplus \ldots \oplus v_r$ is in this intersection then for any 1-linearization $R = k[\phi' v_1, \ldots, \phi' v_r]$ and it is thus clear that $v_1, \ldots, v_r$ is 1-admissible. Now Proposition 7.2 asserts the existence of $w_1, \ldots, w_r$ which is $r+1$ admissible.

Writing

$$f_i = \phi(w_i), \qquad i = 1, \ldots, r$$

we have $\operatorname{Dim}(f_{i_1}, \ldots, f_{i_s}) = r - s$. q.e.d.

8. Bounds for the Betti Numbers

Let k be an infinite field, and let (B, A) satisfy S.A.C. Let $R = k[u_1, \ldots, u_r]$ be graded with $\deg u_i = a_i$, and let $R' = k[u_1, \ldots, u_r, v_{r+1}, \ldots, v_q]$, with each v_i assigned degree 1.

Proposition 8.1 and its proof (including Lemma 8.2) are due to C. Allday.

8.1. **Proposition.** *(Allday). Assume $f_i \in R$, $(1 \leqq i \leqq q)$ are homogeneous with* $\deg f_i = b_i$ *and satisfy*

$$\dim_k R/(f_1, \ldots, f_q) < \infty.$$

Then there exist $g_i \in k[v_{r+1}, \ldots, v_q]$, $1 \leqq i \leqq q$, homogeneous with $\deg g_i = b_i$ *and such that*

$$\dim_k R'/(f_1 + g_1, \ldots, f_q + g_q) < \infty.$$

8.2. **Lemma.** *Let $m \leqq q$ and assume that $g_1, \ldots, g_{m-1}$ have been chosen so that, for $1 \leqq i < m$,*

$$\operatorname{Dim}(f_1 + g_1, \ldots, f_i + g_i) = q - i.$$

Let $P_1, \ldots, P_l$ be the isolated prime ideals of $I = (f_1 + g_1, \ldots, f_{m-1} + g_{m-1})$ and let $L \subset k[v_{r+1}, \ldots, v_q]$ be the space of polynomials of degree b_m. Then, the affine space $f_m + L$ is not contained in the union $\bigcup_1^l P_\nu$.

Proof of Lemma. If $f_m + L \subset \bigcup_1^l P_\nu$, then it follows that $f_m + L \subset P_\nu$ for some ν, say $\nu = 1$. This gives $f_m \in P_1$ and $L \subset P_1$. Since $v_i^{b_m} \in L$ for all i and P_1 is prime, it follows that

(8.3) $\quad f_m, v_{r+1}, \ldots, v_q \quad$ are all in P_1.

Since I is a graded ideal so are all the P_ν [17; Corollary p. 154].

Letting $Q = P_1 \cap R$, it follows from (8.3) that $R'/P_1 \cong R/Q$; whence Q is a graded prime ideal and $\operatorname{Dim} Q = \operatorname{Dim} P_1$.

The elements $f_{m+1}, \ldots, f_q$ generate a proper ideal $\bar{Q}$ in R/Q and so, by Lemma 5.1, $\operatorname{Dim} \bar{Q} \geqq \operatorname{Dim} Q - (q - m)$.

On the other hand by (8.3), $f_i \in Q$ for $1 \leqq i \leqq m$, so

$$\dim_k (R/Q)/\bar{Q} < \infty,$$

whence $\operatorname{Dim} \bar{Q} = 0$. Thus,

$$0 = \operatorname{Dim} \bar{Q} \geqq \operatorname{Dim} Q - (q-m) \geqq \operatorname{Dim} P_1 - (q-m).$$

However, since $\operatorname{Dim} I = q-(m-1)$, Macaulay's theorem [17; Theorem 26, p. 203] asserts that $\operatorname{Dim} P_1 = q-(m-1)$, so

$$0 \geqq \operatorname{Dim} P_1 - (q-m) = q-(m-1)-(q-m) = 1.$$

This contradiction establishes the lemma. q.e.d.

Proof of Proposition 8.1. We construct $g_1, \ldots, g_q$ inductively so that

$$\operatorname{Dim}(f_1+g_1, \ldots, f_i+g_i) = q-i, \qquad \text{for } 1 \leqq i \leqq q.$$

We let $m \leqq q$ and assume that g_i has been chosen for $i<m$.

In the notation of Lemma 8.2, we have

$$f_m + L \not\subset \bigcup_1^l P_\nu.$$

Choose g_m so that $f_m + g_m \notin P_\nu$ for $1 \leqq \nu \leqq l$; then every isolated prime P of $J=(f_1 + g_1, \ldots, f_m+g_m)$ satisfies $P \supsetneqq P_\nu$ for some ν. Hence, [17; (2), p. 193],

$$\operatorname{Dim} P < \operatorname{Dim} P_\nu \leqq \operatorname{Dim} I = q-m+1.$$

Since this holds for each isolated P, $\operatorname{Dim} J \leqq q-m$. But, since J is generated by m elements, $\operatorname{Dim} J \geqq q-m$ by Lemma 5.1. Thus $\operatorname{Dim} J = q-m$ and the proof is complete. q.e.d.

Proof of Proposition 2.8.

(1) Let S be a space of type F with exponents $B=(b_1, \ldots, b_q)$; $A=(a_1, \ldots, a_r)$. Let $B'=(2b_1, \ldots, 2b_q)$ and $A'=(2a_1, \ldots, 2a_q)$ where $a_{r+1} = \ldots = a_q = 1$. Since, by theorem 1, (B, A) satisfies S.A.C. it follows that (B', A') satisfies A.C. By Proposition 3.1 the expression

$$\prod_1^q (1-t^{2b_i}) \Big/ \left[(1-t^2)^{q-r} \prod_1^r (1-t^{2a_i})\right]$$

is a polynomial with non-negative coefficients, whence, a fortiori, so is

$$\prod_1^q (1-t^{2b_i}) \Big/ (1-t)^{q-r} \prod_1^r (1-t^{2a_i}).$$

Since, by (2.2), this latter polynomial has degree $n(S)$, we get (1).

(2) Let $(\Lambda X, d)$ be the minimal model of S and choose bases $x_1, \ldots, x_q$ of X^{odd} and $y_1, \ldots, y_r$ of X^{even} so that $\deg x_i = 2b_i - 1$, $\deg y_i = 2a_i$. According to [9; §5]

there is, a spectral sequence converging to $H(\Lambda X, d)$, and whose E_1 term is of the form $(\Lambda X, \bar{d})$ with

$$\bar{d}x_i \in \Lambda(y_1, \dots, y_r) \quad \text{and} \quad \bar{d}y_j = 0 \text{ for all } i, j.$$

It is thus sufficient to prove that, for $0 \leqq m \leqq n(S)$,

$$\dim H^m(\Lambda X, \bar{d}) \leqq c_m.$$

By [9; Proposition 1, §6], $\dim H(\Lambda X, \bar{d}) < \infty$, so, writing $\bar{d}x_i = f_i$ we find that

$$\dim \Lambda(y_1, \dots, y_r)/(f_1, \dots, f_q) < \infty.$$

By (8.1), we may choose variables $v_{r+1}, \dots, v_q$ of degree 2 and homogeneous polynomials $g_1, \dots, g_q$ in $\Lambda(v_{r+1}, \dots, v_q)$ so that $\deg g_i = 2b_i$ and

(8.4) $\dim \Lambda(y_1, \dots, y_r, v_{r+1}, \dots, v_q)/(f_1 + g_1, \dots, f_q + g_q) < \infty.$

Let V be the span of the v_i and, in $\Lambda X \otimes \Lambda V$, define a differential D by:

$$Dx_i = f_i + g_i, \; Dy_j = 0, \; Dv_l = 0, \qquad \text{for all } i, j, l.$$

By (8.4), $\dim H(\Lambda X \otimes \Lambda V, D) < \infty$ and hence,

$$\Sigma \dim H^m(\Lambda X \otimes \Lambda V)\, t^m = \Pi(1 - t^{2b_i})/(1 - t^2)^{q-r}\, \Pi(1 - t^{2a_i}).$$

Finally, let W be a vector space with basis $w_{r+1}, \dots, w_q$, all of degree 1 and extend D to $\Lambda W \otimes \Lambda X \otimes \Lambda V$ by setting $Dw_i = v_i$. Then $H(\Lambda X, \bar{d}) \cong H(\Lambda W \otimes \Lambda X \otimes \Lambda V, D)$. Since there is, converging to $H(\Lambda W \otimes \Lambda X \otimes \Lambda V, D)$, a spectral sequence whose E_1 term is given by $\Lambda W \otimes H(\Lambda X \otimes \Lambda V, D)$, it follows that

(8.5) $\dim H^m(\Lambda X, \bar{d}) \leqq \sum_s \dim \Lambda^s W \dim H^{m-s}(\Lambda X \otimes \Lambda V, D).$

Now,

$$\begin{aligned}
&\sum_{m,s} \dim \Lambda^s W \dim H^{m-s}(\Lambda X \otimes \Lambda V, D)\, t^m \\
&\quad = \Big(\sum_s \dim \Lambda^s W \cdot t^s\Big)\Big(\sum_u \dim H^u(\Lambda X \otimes \Lambda V, D)\, t^u\Big) \\
&\quad = (1+t)^{q-r} \prod_1^q (1 - t^{2b_i})/(1 - t^2)^{q-r} \prod_1^r (1 - t^{2a_i}) \\
&\quad = \prod_1^q (1 - t^{2b_i})/(1 - t)^{q-r} \prod_1^r (1 - t^{2a_i}).
\end{aligned}$$

Hence,

$$\sum_s \dim \Lambda^s W \dim H^{m-s}(\Lambda X \otimes \Lambda V, D) = c_m.$$

This, together with (8.5), completes the proof. q.e.d.

References

1. Allday, C., Halperin, S.: Lie group actions on spaces of finite rank, Quart. J. Math. (Oxford) **29**, 63–76 (1978)
2. Barge, J.: Structures différentiables sur les types d'homotopie rationelle simplement connexes, Thèse, Univ. de Paris Sud (Orsay), 1975
3. Bousfield, A.K., Gugenheim, V.K.A.M.: On PL de Rham theory and rational homotopy type, Memoirs A.M.S. 179, 1976
4. Cartan, H.: La transgression dans un groupe de Lie et dans un espace fibré principal. Colloque de Topologie (espaces fibrés) pp. 57–71. Bruxelles (1950), Thone, Liège, Paris: Masson 1951
5. Greub, W., Halperin, S., Vanstone, J.R.: Connections, Curvature and Cohomology, Vol. III, New York: Academic Press 1976
6. Grosswald, E.: Reducible rational fractions of the type of Gaussian polynomials with only non-negative coefficients, Canadian Math. Bull. in press (1979)
7. Hall, P.: On representatives of subsets, J. London Math. Soc. **10**, 26–30 (1935)
8. Halperin, S.: Lectures on minimal models, Publ. Internes de l'Univ. de Lille I, 111, 1977
9. Halperin, S.: Finiteness in the minimal models of Sullivan, Trans. Amer. Math. Soc. **230**, 173–199 (1977)
10. Halperin, S.: Rational fibrations, minimal models, Trans. Amer. Math. Soc. in press (1979)
11. Koszul, J.L.: Sur un type d'algebres différentielles en rapport avec la transgression, Colloque de Topologie (espaces fibrés) pp. 73–81. Bruxelles (1950), Thone, Liège; Paris: Masson MR 13, 109, 1951
12. Lehmann, D.: Théorie homotopique des formes différentielles, Astérisque, **45**, 1–102 (1977)
13. Quillen, D.G.: The spectrum of an equivariant cohomology ring, I, II, Ann. of Math. **94**, 549–602 (1971)
14. Reich, D.: On certain polynomials of Gaussian type, (preprint)
15. Sullivan, D.: Infinitesimal computations in topology, Publ. de l'I.H.E.S., **47**, 269–331 (1978)
16. Zariski, O., Samuel, P.: Commutative algebra, Vol. I, Princeton, N.J.: Van Nostrand 1958
17. Zariski, O., Samuel, P.: Commutative algebra, Vol. II, Princeton, N.J.: Van Nostrand 1960

Received December 11, 1978/Revised April 11, 1979

Inventiones math. 53, 135–164 (1979)

Inventiones mathematicae

On Solvable Number Fields

Jürgen Neukirch

Universität Regensburg, Fachbereich Mathematik, Universitätsstraße 31, D-8400 Regensburg, Federal Republic of Germany

Introduction

By a celebrated theorem of Šafarevič every finite solvable group G can be realized as the Galois group of a Galois extension of any given finite algebraic number field k. Šafarevič's proof of this theorem however is long and difficult. It was our original intention to give a short and more conceptual proof of the theorem in the case of groups of odd order. During our investigations, however, it turned out that under the odd order assumption the methods which we were following were of a very general nature. They finally led to a general theorem on the absolute Galois group of a finite algebraic number field; this theorem now stands in the center of our work. It yields many more and much sharper results than just the existence of Galois extensions with a given Galois group of odd order.

As a first consequence of this main theorem we obtain the result that also every profinite separable group G of finite odd exponent occurs as a Galois group over any given finite algebraic number field. The condition that the exponent of G is finite, i.e. that the elements of G have a finite bounded order, is obviously indispensable, since for example $\mathbb{Z}_p \times \mathbb{Z}_p$ cannot occur as a Galois group over the field $\mathbb{Q}$ of rational numbers, $\mathbb{Z}_p$ being the group of p-adic integers.

Almost all of the applications of the main theorem require the assumption, that the exponents of the profinite groups are prime to the number of roots of unity in the number fields considered. In order to explain some further results let us assume for simplicity and without further mention that the number fields which we consider are real, i.e. admit a real imbedding, since then the assertions are valid for all profinite groups of finite odd exponent.

As a next consequence of the main theorem we have the following result of Grunwald type. Let G be a profinite separable group of finite odd exponent. Let S be a finite set of primes of the number field k and for each $\mathfrak{p} \in S$ let $K_\mathfrak{p} | k_\mathfrak{p}$ be a Galois extension of the completion $k_\mathfrak{p}$ of k, such that the Galois groups $\mathrm{Gal}(K_\mathfrak{p} | k_\mathfrak{p})$ can be imbedded into G. Then there exists a global Galois extension $K | k$, which on the one hand has the given group G as Galois group and which on the other hand has the

0020-9910/79/0053/0135/$06.00

given extensions $K_\mathfrak{p}|k_\mathfrak{p}$ as completions at the primes $\mathfrak{p}\in S$. For cyclic groups G of odd order this is the theorem of Grunwald – Hasse – Wang.

A strong sharpening of the inverse problem of Galois theory is the so called *imbedding problem*: A given Galois extension $K|k$ with Galois group Γ shall be imbedded into a larger one $N|k$, such that the canonical homomorphism $\mathrm{Gal}(N|k)\to\mathrm{Gal}(K|k)$ realizes an abstractly given group extension $G\to\Gamma$. We shall give in a sense a complete answer to this question in the case where the kernel of the homomorphism $G\to\Gamma$ is a profinite separable group of finite odd exponent. The main theorem implies a local-global-principle for this imbedding problem, by which the solvability question is reduced to a purely group theoretical "word problem", granted the fact that the structure of the absolute Galois groups over the $\mathfrak{p}$-adic number fields $k_\mathfrak{p}$ is known by generators and relations.

From this one can obtain many special results as, for instance, that the imbedding problem is always solvable if the given extension $K|k$ is unramified. It has also always a solution if the group extension $G\xrightarrow{f}\Gamma$ splits. This last result has been proven by Scholz in the case that the kernel of f is a finite abelian group and by Šafarevič in the case that the kernel is nilpotent.

Since the theory presented here is dominated to a large part by cohomological methods, it may appear that the original ideas are hidden behind more abstract mechanisms. We therefore want now to explain the procedure of our construction of solvable number fields and clarify the principles which underly it.

The fields are constructed in abelian steps. Therefore we have to study the imbedding problem with abelian kernel, i.e. we have a given Galois extension $K|k$ with Galois group Γ and a group extension $G\to\Gamma$ with a finite abelian kernel A. Then A is a Γ-module, and it is sufficient to study the case where A is a *simple* Γ-module. The problem is to find a Galois extension $N|k$ containing K such that the group extension $\mathrm{Gal}(N|k)\to\mathrm{Gal}(K|k)$ is isomorphic to the given one $G\to\Gamma$. This problem is in general unsolvable.

If $\mathfrak{p}$ runs through the primes of k then we can consider the completions $K_\mathfrak{p}|k_\mathfrak{p}$ of $K|k$ with the Galois groups $\Gamma_\mathfrak{p}\subseteq\Gamma$. Taking the pre-image $G_\mathfrak{p}$ of $\Gamma_\mathfrak{p}$ in G we obtain a group extension $G_\mathfrak{p}\to\Gamma_\mathfrak{p}$ and thus for each $\mathfrak{p}$ a *local imbedding problem*. If $N|k$ is a solution of the global imbedding problem then its completions $N_\mathfrak{p}|k_\mathfrak{p}$ are solutions of the local ones (it may be that the Galois group of $N_\mathfrak{p}|k_\mathfrak{p}$ is only a subgroup of $G_\mathfrak{p}$, but this is unimportant). This observation leads to the

First Principle. In order that the global imbedding problem has a solution it is necessary that all the local ones have a solution.

We have therefore first to study the question, as to when the local imbedding problem is solvable. Clearly, it is necessary to find a criterion as simple and as general as possible. It turns out that this forces one to admit only cyclic extensions $K_\mathfrak{p}|k_\mathfrak{p}$. If $K_\mathfrak{p}|k_\mathfrak{p}$ is unramified then the imbedding problem has always a solution. If however $K_\mathfrak{p}|k_\mathfrak{p}$ is ramified, then the cyclicity alone is not sufficient. The typical situation is this. Let π be a prime element of $k_\mathfrak{p}$ and $K_\mathfrak{p}=k_\mathfrak{p}(\sqrt[n]{\pi})$. In order to imbed $K_\mathfrak{p}|k_\mathfrak{p}$ into a larger cyclic extension of degree m (prime to $\mathfrak{p}$) it is necessary and sufficient that $k_\mathfrak{p}$ contains the m-th roots of unity, or in other words, that the prime $\mathfrak{p}$ of k splits completely in the field $\Omega=k(\zeta_m)$ of m-th roots of unity. Then the field

$k_\mathfrak{p}(\sqrt[m]{\pi})$ can be taken as a solution of the imbedding problem. These local studies lead to the following principles necessary for the global construction.

Second Principle. One has to restrict oneself to extensions $K|k$ which are *locally cyclic*, i.e. for which the completions are cyclic.

Third Principle. The extension $K|k$ must have the property that the primes $\mathfrak{p}$ which ramify in K are totally split in a given field $\Omega|k$ of roots of unity.

It will turn out that the true state of affairs is not quite so restrictive as these principles indicate, since at finitely many given primes one has an arbitrary freedom. They play however a leading role in the construction.

We now ask whether these conditions are also sufficient for the existence of a solution $N|k$ of our global imbedding problem, i.e. whether there exists a local-global-principle for it. In general such a local-global-principle is not available. However we are in our situation very lucky. Since the Γ-module A is simple, it is annihilated by a prime number p and we have

Fourth Principle. If A is a simple Γ-module annihilated by p and if K does *not* contain the group μ_p of p-th roots of unity, then the global imbedding problem is solvable if and only if all the local ones are.

This local-global-principle was pointed out to me by O. Neumann. Since every field contains the second roots of unity, it seems that at this point we are definitely restricted to groups of odd order. But this is actually not the case since the local-global-principle as formulated here serves only to simplify our arguments. The condition $\mu_p \nsubseteq K$ will however become very important later.

We have obtained the result that the imbedding problem has a solution $N|k$ if the extension $K|k$ is locally cyclic, if the ramified primes split completely in a field Ω of roots of unity and if $\mu_p \nsubseteq K$. But this does not mean that we are finished. On the contrary, the essential difficulties begin at this point. Namely, in the next step we have to imbed the extension $N|k$ into a larger one, i.e. we have to solve a further imbedding problem, in which the role of the extension $K|k$ is taken over by $N|k$. Therefore we shall have to require from $N|k$ that it is locally cyclic, that its ramification primes split completely in the field Ω and that $\mu_p \nsubseteq N$. This requirement strengthens very much the original imbedding problem and we come to the

Fifth Principle. The solution $N|k$ of the imbedding problem must remain locally cyclic.

Sixth Principle. The primes which ramify in $N|k$ must split completely in Ω.

The condition $\mu_p \nsubseteq N$ causes no essential problem. In order to explain how to achieve the local cyclicity let us go from the simple Γ-module A to the induced one $M_\Gamma(\mathbb{Z}/p)$ of which A is a homomorphic image. To solve the imbedding problem with this new kernel means only to find a cyclic extension $L|K$ of degree p of which the conjugates over k are independent. The composite N of these conjugates will then be a solution of the imbedding problem.

The finite set of primes which are ramified in K can be handled separately. The problem is to throw the primes $\mathfrak{p}$ which are unramified in K but ramified in N into the set of primes which split completely in Ω and to keep the local extensions $N_\mathfrak{p}|k_\mathfrak{p}$ at such primes cyclic. This is achieved as follows. One constructs the cyclic extension $L|K$ in such a way that the primes $\mathfrak{p}$ of k which ramify in L split completely not only in Ω but also in K. Moreover one forces them to behave in the following way. If $\mathfrak{P}$ is a prime of K lying over $\mathfrak{p}$ which ramifies in L then the primes $\mathfrak{P}'$ conjugate to $\mathfrak{P}$ over $\mathfrak{p}$ are split in L. This means in other words that $\mathfrak{P}$ splits completely in the fields conjugate to L over k, so that the completion $N_\mathfrak{p}|k_\mathfrak{p}$ of the composite $N|k$ becomes cyclic. The conjugation problem can really be solved, although it causes the main technical difficulties. The problem of forcing the ramified primes to split completely in the extension ΩK is of a more severe nature. If K contains the p-th roots of unity this is actually impossible and one obtains non vanishing obstructions for the solvability of the sharpened imbedding problem. The principal reason for the relative simplicity of our theory in contrast to Šafarevič's procedure is the observation that there is no obstruction if $\mu_p \nsubseteq K$ and that under this assumption also the splitting problem can be solved. On the other hand it is this point which restricts us to the groups of odd order since every field contains the second roots of unity.

After these considerations we find an extension $N|k$ solving the imbedding problem and having all the properties which are necessary to be imbedded into a larger field etc. Let us finally mention, why this successive construction must stop and can only lead to extensions with a Galois group of bounded exponent. Let $\Omega|k$ be the field of n-th roots of unity. If then we have come to a point where the completion $N_\mathfrak{p}|k_\mathfrak{p}$ of the constructed extension $N|k$ is totally ramified of degree n, then we cannot imbed it anymore into a cyclic extension of larger degree since we cannot assume that the ground field $k_\mathfrak{p}$ contains more than the n-th roots of unity. Therefore at this stage the procedure stops.

We have only explained the essential principles and ideas underlying the theory which we shall develop. Our actual methods, as mentioned before, will be more abstract. Although inspired by the work of Šafarevič they differ to a large degree from Šafarevič's methods. It is possible that they might also be useful for a future simplification of the inverse problem of Galois theory in the case of solvable groups of even order. But the generality of the results presented here will certainly not carry over to the even order case.

§ 1. Cohomology of Simple Galois Modules

Let K be a field and $p \neq \mathrm{char}(K)$ a prime number which is fixed once and for all in this paragraph. Let μ_p be the group of p-th roots of unity and let $\Delta = \mathrm{Gal}(\mathscr{K}|K)$ be the Galois group of the field $\mathscr{K} = K(\mu_p)$. We first recall some facts about Δ-decompositions.

We have the canonical character

$$\theta\colon \Delta \to (\mathbb{Z}/p)^*,$$

given by $\delta\zeta = \zeta^{\theta(\delta)}$ for $\delta \in \Delta$ and $\zeta \in \mu_p$. The idempotents

$$e_i = \frac{1}{d} \sum_{\delta \in \Delta} \theta(\delta)^{-i} \delta, \qquad d = \# \Delta, \ i \in \mathbb{Z},$$

in the group ring $\mathbb{Z}/p[\Delta]$ yield for any $\mathbb{Z}/p[\Delta]$-module A a decomposition

$$A = \bigoplus_{i=1}^{d-1} e_i A,$$

and $e_i A$ is the maximal submodule of A on which the elements $\delta \in \Delta$ act as multiplication by $\theta(\delta)^i$.

For $n \in \mathbb{Z}$ the n-fold *Tate-twist* of the $\mathbb{Z}/p[\Delta]$-module A is the $\mathbb{Z}/p[\Delta]$-module $A(n) = A \otimes \mu_p^{\otimes n}$, where $\mu_p^{\otimes n} = \mathbb{Z}/p$ if $n=0$ and $\mu_p^{\otimes n} = \mathrm{Hom}(\mu_p^{\otimes(-n)}, \mathbb{Z}/p)$ if $n<0$. As an abelian group $A(n)$ is isomorphic to A, but the Δ-action $\delta\colon a \mapsto \delta a$ is changed into the Δ-action $\delta\colon a \mapsto \theta(\delta)^n \delta a$. We have

$$A(n)^{\Delta} = (e_{-n} A)(n).$$

Almost all of our considerations in this paper rely on the assumption, that the field K does *not* contain the p-th roots of unity. The reason for the simplicity and generality of our theory of solvable number fields under this assumption relies on the following observation.

Let $\mathscr{K}'|\mathscr{K}$ be the maximal abelian extension of exponent p of the field $\mathscr{K} = K(\mu_p)$. Then the Galois group $\mathrm{Gal}(\mathscr{K}'|\mathscr{K})$ is a $\mathbb{Z}/p[\Delta]$-module and we have

Lemma 1. *Assume that $\mu_p \nsubseteq K$ and let $\Omega|K$ be an abelian extension. Then the group $e_1 \,\mathrm{Gal}(\mathscr{K}'|\mathscr{K}) \subseteq \mathrm{Gal}(\mathscr{K}'|\mathscr{K})$ acts trivially on the field $\Omega \cap \mathscr{K}'$.*

Proof. We can assume that $\mu_p \subseteq \Omega$, i.e. $K \subseteq \mathscr{K} \subseteq \Omega$. Since $\Omega|K$ is abelian the group $\Delta = \mathrm{Gal}(\mathscr{K}|K)$ acts trivially on $\mathrm{Gal}(\Omega \cap \mathscr{K}'|\mathscr{K})$. If $\tau \in e_1 \,\mathrm{Gal}(\mathscr{K}'|\mathscr{K})$ and $\bar{\tau} = \tau|_{\Omega \cap \mathscr{K}'}$, then $\bar{\tau}^{\delta} = \bar{\tau}^{\theta(\delta)} = \bar{\tau}$. Since $\mu_p \nsubseteq K$ the group Δ is non-trivial, i.e. $\theta(\delta)$ is not always 1 and hence $\bar{\tau} = 1$.

If G_K is the absolute Galois group of the field K and A is a G_K-module, then we denote by $H^q(K, A)$ the Galois cohomology $H^q(G_K, A)$. We call a cohomology class $x \in H^q(K, A)$ *cyclic*, if it is split by a cyclic extension of K, i.e. if there exists a cyclic extension $L|K$ such that x lies in the kernel of the restriction map

$$H^q(K, A) \to H^q(L, A).$$

Let now K be a finite algebraic number field with completions $K_{\mathfrak{p}}$ at the primes $\mathfrak{p}$ of K. We shall consider the homomorphisms

$$H^q(K, A) \to \prod_{\mathfrak{p}} H^q(K_{\mathfrak{p}}, A),$$

where the restricted product is taken with respect to the unramified part $H^q_{nr}(K_{\mathfrak{p}}, A)$ of $H^q(K_{\mathfrak{p}}, A)$. If $x \in H^q(K, A)$ then $x_{\mathfrak{p}}$ denotes the image of x in $H^q(K_{\mathfrak{p}}, A)$. We call $x_{\mathfrak{p}}$ *ramified*, if $x_{\mathfrak{p}} \notin H^q_{nr}(K_{\mathfrak{p}}, A)$.

Main Lemma. *Let $K|k$ be a Galois extension and $\Omega|K$ an abelian extension. Assume that $\mu_p \nsubseteq K$.*

Let A be a trivial finite G_K-module such that $pA=0$. Let S be a finite set of primes of K and let $y_\mathfrak{p} \in H^1(K_\mathfrak{p}, A)$ for $\mathfrak{p} \in S$. Then there exists an element $x \in H^1(K, A)$ such that

(i) *$x_\mathfrak{p} = y_\mathfrak{p}$ for $\mathfrak{p} \in S$.*

(ii) *If $\mathfrak{p} \notin S$ then $x_\mathfrak{p}$ is cyclic, and if $x_\mathfrak{p}$ is ramified, then the prime $\mathfrak{p}_0 = \mathfrak{p} \cap k$ of k lying under $\mathfrak{p}$ splits completely in Ω and $x_{\mathfrak{p}'} = 0$ for all primes $\mathfrak{p}' | \mathfrak{p}_0$ of K different from $\mathfrak{p}$.*

Proof. We first prove the lemma in the case $A = \mathbb{Z}/p$. Let $\mathscr{K} = K(\mu_p)$ and $\Delta = \mathrm{Gal}(\mathscr{K}|K)$. For every prime $\mathfrak{p}$ of K we set

$$I_\mathfrak{p} = \prod_{\mathfrak{P}|\mathfrak{p}} \mathscr{K}_\mathfrak{P}^* / \mathscr{K}_\mathfrak{P}^{*p} \quad \text{and} \quad I_\mathfrak{p}^0 = \prod_{\mathfrak{P}|\mathfrak{p}} U_\mathfrak{P} \mathscr{K}_\mathfrak{P}^{*p} / \mathscr{K}_\mathfrak{P}^{*p},$$

where $\mathfrak{P}$ runs through the primes of $\mathscr{K}$ above $\mathfrak{p}$ and where $U_\mathfrak{P}$ is the group of units in $\mathscr{K}_\mathfrak{P}$ ($U_\mathfrak{P} = \mathscr{K}_\mathfrak{P}^*$ if $\mathfrak{P}$ is infinite). Let $\mathscr{K}'|\mathscr{K}$ be the maximal abelian extension of exponent p of $\mathscr{K}$. Then by class field theory we have an exact sequence of Δ-modules

$$1 \to \mathscr{K}^*/\mathscr{K}^{*p} \to \prod_\mathfrak{p} I_\mathfrak{p} \xrightarrow{(\ ,\mathscr{K}'|\mathscr{K})} \mathrm{Gal}(\mathscr{K}'|\mathscr{K}) \to 1.$$

The map $(\ ,\mathscr{K}'|\mathscr{K})$ is given by the converging product

$$(a, \mathscr{K}'|\mathscr{K}) = \prod_\mathfrak{P} (a_\mathfrak{P}, \mathscr{K}'_\mathfrak{P}|\mathscr{K}_\mathfrak{P})$$

for $a = (a_\mathfrak{P}) \in \prod_\mathfrak{P} \mathscr{K}_\mathfrak{P}^*/\mathscr{K}_\mathfrak{P}^{*p}$, where $\mathscr{K}'_\mathfrak{P}|\mathscr{K}_\mathfrak{P}$ is the maximal abelian extension of exponent p of $\mathscr{K}_\mathfrak{P}$ and

$$(\ ,\mathscr{K}'_\mathfrak{P}|\mathscr{K}_\mathfrak{P}): \mathscr{K}_\mathfrak{P}^*/\mathscr{K}_\mathfrak{P}^{*p} \to \mathrm{Gal}(\mathscr{K}'_\mathfrak{P}|\mathscr{K}_\mathfrak{P}) \subseteq \mathrm{Gal}(\mathscr{K}'|\mathscr{K})$$

is the local norm residue symbol. If $\alpha \in \mathscr{K}^*/\mathscr{K}^{*p}$ then $\alpha_\mathfrak{p}$ denotes the image of α in $I_\mathfrak{p}$.

Now the group $H^1(\mathscr{K}, \mathbb{Z}/p)$ is a Δ-module and we have the canonical isomorphisms

$$H^1(K, \mathbb{Z}/p)(1) \cong H^1(\mathscr{K}, \mathbb{Z}/p)^\Delta(1) \cong e_1 H^1(\mathscr{K}, \mu_p) \cong e_1 \mathscr{K}^*/\mathscr{K}^{*p}.$$

The field theoretical meaning is this: Let ζ be a primitive p-th root of unity. Let $x \in H^1(K, \mathbb{Z}/p)$ and let $\alpha \in e_1 \mathscr{K}^*/\mathscr{K}^{*p} \subset \mathscr{K}^*/\mathscr{K}^{*p}$ be the element corresponding to $x \otimes \zeta$. If $K_x|K$ is the cyclic extension of K defined by x then

$$\mathscr{K}(\sqrt[p]{\alpha}) = \mathscr{K} \cdot K_x.$$

For every prime $\mathfrak{p}$ of K the product $\prod_{\mathfrak{P}|\mathfrak{p}} H^1(\mathscr{K}_\mathfrak{P}, \mathbb{Z}/p)$ is the induced Δ-module $M_\Delta^{\Delta_\mathfrak{p}}(H^1(\mathscr{K}_\mathfrak{P}, \mathbb{Z}/p))$, where $\mathfrak{P}$ is a selected prime of $\mathscr{K}$ above $\mathfrak{p}$ and $\Delta_\mathfrak{p} \subseteq \Delta$ is the corresponding decomposition group. From this we deduce

$$\begin{aligned} H^1(K_\mathfrak{p}, \mathbb{Z}/p)(1) &\cong \big(\prod_{\mathfrak{P}|\mathfrak{p}} H^1(\mathcal{K}_\mathfrak{P}, \mathbb{Z}/p)\big)^\Delta(1) \\ &\cong e_1\big(\prod_{\mathfrak{P}|\mathfrak{p}} H^1(\mathcal{K}_\mathfrak{P}, \mu_p)\big) \cong e_1\big(\prod_{\mathfrak{P}|\mathfrak{p}} \mathcal{K}_\mathfrak{P}^*/\mathcal{K}_\mathfrak{P}^{*p}\big). \end{aligned}$$

If we choose a primitive p-th root of unity ζ then the correspondence $y \mapsto y \otimes \zeta$ yields isomorphisms

$$H^1(K, \mathbb{Z}/p) \cong H^1(K, \mathbb{Z}/p)(1), \qquad H^1(K_\mathfrak{p}, \mathbb{Z}/p) \cong H^1(K_\mathfrak{p}, \mathbb{Z}/p)(1).$$

All these considerations together yield a commutative diagram

$$\begin{tikzcd}
1 \arrow[r] & \mathcal{K}^*/\mathcal{K}^{*p} \arrow[r] & \prod_\mathfrak{p} I_\mathfrak{p} \arrow[r, "{(\ ,\mathcal{K}'|\mathcal{K})}"] & \mathrm{Gal}(\mathcal{K}'|\mathcal{K}) \arrow[r] & 1 \\
1 \arrow[r] & e_1\mathcal{K}^*/\mathcal{K}^{*p} \arrow[r] \arrow[u, hook] & \prod_\mathfrak{p} e_1 I_\mathfrak{p} \arrow[r] \arrow[u, hook] & e_1\mathrm{Gal}(\mathcal{K}'|\mathcal{K}) \arrow[r] \arrow[u, hook] & 1 \\
1 \arrow[r] & H^1(K, \mathbb{Z}/p) \arrow[r] \arrow[u, "\wr"] & \prod_\mathfrak{p} H^1(K_\mathfrak{p}, \mathbb{Z}/p) \arrow[u, "\wr"] & &
\end{tikzcd}$$

With these preparations we now proceed to the proof. We can obviously assume that $\Omega \supseteq \mathcal{K}$ and that the set S is closed under conjugation over k, that it contains all primes above p, all infinite primes and all primes $\mathfrak{p}$ of K such that $\mathfrak{p} \cap k$ is ramified in Ω. Let $\beta_\mathfrak{p}$ be the image of $y_\mathfrak{p} \in H^1(K_\mathfrak{p}, \mathbb{Z}/p)$ in $e_1 I_\mathfrak{p} \subset I_\mathfrak{p}$ for $\mathfrak{p} \in S$. Then

$$\tau = \prod_{\mathfrak{p} \in S} (\beta_\mathfrak{p}, \mathcal{K}'|\mathcal{K}) \in e_1 \mathrm{Gal}(\mathcal{K}'|\mathcal{K}).$$

By Lemma 1 τ is the identity on the field $\Omega \cap \mathcal{K}'$, which is the maximal subextension of exponent p of $\Omega|\mathcal{K}$.

Now by the theorem which we shall prove in the appendix of this paper we find an element $\gamma \in \mathcal{K}^*/\mathcal{K}^{*p}$ with the properties

(1) $\gamma_\mathfrak{p} = \beta_\mathfrak{p}$ for $\mathfrak{p} \in S$.
(2) If $\gamma_\mathfrak{p} \notin I_\mathfrak{p}^0$ for $\mathfrak{p} \notin S$ then $\mathfrak{p} \cap k$ splits completely in Ω and $\gamma_{\mathfrak{p}'} = 1$ for the primes $\mathfrak{p}'|\mathfrak{p} \cap k$ of K different from $\mathfrak{p}$.

The element $\alpha = e_1\gamma \in e_1\mathcal{K}^*/\mathcal{K}^{*p}$ satisfies the same conditions since $\beta_\mathfrak{p} \in e_1 I_\mathfrak{p}$ and therefore $e_1\beta_\mathfrak{p} = \beta_\mathfrak{p}$.

Let x be the element in $H^1(K, \mathbb{Z}/p)$ corresponding to $\alpha \in e_1\mathcal{K}^*/\mathcal{K}^{*p}$. Then the element $\alpha_\mathfrak{p} \in e_1 I_\mathfrak{p}$ corresponds to $x_\mathfrak{p}$ and we obtain

(3) $x_\mathfrak{p} = y_\mathfrak{p}$ for $\mathfrak{p} \in S$.
(4) If $x_\mathfrak{p}$ is ramified for $\mathfrak{p} \notin S$ then $\alpha_\mathfrak{p} \notin e_1 I_\mathfrak{p}^0$, i.e. $\mathfrak{p}_0 = \mathfrak{p} \cap k$ splits completely in Ω and $\alpha_{\mathfrak{p}'} = 1$, i.e. $x_{\mathfrak{p}'} = 0$ for $\mathfrak{p}'|\mathfrak{p}_0$, $\mathfrak{p}' \neq \mathfrak{p}$.

This proves the main lemma in the case $A = \mathbb{Z}/p$.

The general case will be proven by induction over $\dim_{\mathbb{Z}/p} A$. Again, we can assume that S is closed under conjugation over k. Let $A = A' \oplus \mathbb{Z}/p$. For each $x \in H^1(K, A)$ let $x = x' + x''$ be the decomposition of x into the components $x' \in H^1(K, A')$, $x'' \in H^1(K, \mathbb{Z}/p)$, and similarly $x_\mathfrak{p} = x'_\mathfrak{p} + x''_\mathfrak{p}$ for $x_\mathfrak{p} \in H^1(K_\mathfrak{p}, A)$. By induction we find an element $x' \in H^1(K, A')$ such that

(5) $x'_\mathfrak{p} = y'_\mathfrak{p}$ for $\mathfrak{p} \in S$,

(6) If $\mathfrak{p} \notin S$ then $x'_\mathfrak{p}$ is cyclic, and if $x'_\mathfrak{p}$ is ramified then $\mathfrak{p}_0 = \mathfrak{p} \cap k$ splits completely in Ω and $x'_{\mathfrak{p}'} = 0$ for $\mathfrak{p}' | \mathfrak{p}_0$, $\mathfrak{p}' \neq \mathfrak{p}$.

Let $K'|K$ be the abelian extension defined by the homomorphism $x'\colon G_K \to A'$, and set $\Omega' = K'\Omega$. Let V be a finite set of primes of K which is closed under conjugation over k and contains the set $S \cup \{\mathfrak{p} \,|\, x'_\mathfrak{p} \text{ is ramified}\}$. By what we have proved before we find an element $x'' \in H^1(K, \mathbb{Z}/p)$ such that

(7) $x''_\mathfrak{p} = y''_\mathfrak{p}$ for $\mathfrak{p} \in S$, $x''_\mathfrak{p} = 0$ for $\mathfrak{p} \in V - S$,

(8) If $x''_\mathfrak{p}$ is ramified for $\mathfrak{p} \notin V$, then $\mathfrak{p}_0 = \mathfrak{p} \cap k$ splits completely in Ω' and $x''_{\mathfrak{p}'} = 0$ for $\mathfrak{p}' | \mathfrak{p}_0$, $\mathfrak{p}' \neq \mathfrak{p}$.

We now show that the element

$$x = x' + x'' \in H^1(K, A)$$

satisfies the conditions of the lemma. For $\mathfrak{p} \in S$ we have

$$x_\mathfrak{p} = x'_\mathfrak{p} + x''_\mathfrak{p} = y'_\mathfrak{p} + y''_\mathfrak{p} = y_\mathfrak{p}.$$

For $\mathfrak{p} \in V - S$ we have

$$x_\mathfrak{p} = x'_\mathfrak{p} + x''_\mathfrak{p} = x'_\mathfrak{p}.$$

Therefore $x_\mathfrak{p}$ is cyclic and if $x_\mathfrak{p} = x'_\mathfrak{p}$ is ramified then $\mathfrak{p}_0$ splits completely in Ω and $x_{\mathfrak{p}'} = x'_{\mathfrak{p}'} + x''_{\mathfrak{p}'} = 0$ for $\mathfrak{p}' | \mathfrak{p}_0$, $\mathfrak{p}' \neq \mathfrak{p}$. Let finally $\mathfrak{p} \notin V$. If $x_\mathfrak{p}$ is unramified then it is cyclic. Let $x_\mathfrak{p}$ be ramified. Since $x_\mathfrak{p} = x'_\mathfrak{p} + x''_\mathfrak{p}$ and $x'_\mathfrak{p}$ is unramified, $x''_\mathfrak{p}$ must be ramified. This means that $\mathfrak{p}_0$ splits completely in Ω'. In particular it splits completely in K'. From this we obtain $x'_{\mathfrak{p}'} = 0$ for all $\mathfrak{p}' | \mathfrak{p}_0$, since, by definition of K', x' becomes zero in $H^1(K', A')$ and thus $x'_{\mathfrak{p}'}$ is zero in $H^1(K'_{\mathfrak{P}'}, A') = H^1(K_{\mathfrak{p}'}, A')$, $\mathfrak{P}'$ being a prime of K' above $\mathfrak{p}'$. Therefore $x_\mathfrak{p} = x''_\mathfrak{p}$, i.e. $x_\mathfrak{p}$ is cyclic and $x_{\mathfrak{p}'} = x''_{\mathfrak{p}'} = 0$ for $\mathfrak{p}' | \mathfrak{p}_0$, $\mathfrak{p}' \neq \mathfrak{p}$. This proves the main lemma in the general case.

If A is a finite G_k-module we denote by $k(A)|k$ the smallest extension of k over which A becomes a trivial Galois module. In the sequel we will always assume, that A is annihilated by the prime number p, i.e. $pA = 0$. By A' we denote the dual G_k-module $A' = \operatorname{Hom}(A, \mu_p)$.

Lemma 2. *Let A be a simple G_k-module and $K|k$ a Galois extension such that $k(A) \subseteq K$ but $\mu_p \nsubseteq K$.*

Let P be a set of primes of k containing almost all primes which split completely in $K(\mu_p)$. Then the homomorphism

$$H^1(k, A') \to \prod_{\mathfrak{p} \in P} H^1(k_\mathfrak{p}, A')$$

is injective.

Proof. Let $\mathcal{K}=K(\mu_p)$ and $\mathcal{G}=\mathrm{Gal}(\mathcal{K}|k)$, $G=\mathrm{Gal}(K|k)$ and $\Delta=\mathrm{Gal}(\mathcal{K}|K)$. We first show, that

$$H^i(\mathcal{G},A')=0 \qquad \text{for all } i.$$

Since the order of Δ is prime to p we have $H^j(\Delta,A')=0$ for $j>0$. Therefore the Hochschild-Serre spectral sequence

$$H^i(G,H^j(\Delta,A')) \Rightarrow H^{i+j}(\mathcal{G},A')$$

degenerates and yields isomorphisms

$$H^i(G,A'^{\Delta})\cong H^i(\mathcal{G},A').$$

Now A' is a simple $\mathcal{G}$-module. In fact, if B is a $\mathcal{G}$-submodule of A', then the orthogonal complement $B^{\perp}$ under the canonical pairing $A\times A'\to\mu_p$ is a $\mathcal{G}$-submodule of A and since A is simple we have either $B^{\perp}=A$ or $B^{\perp}=0$, i.e. either $B=0$ or $B=A'$.

On the other hand Δ acts non-trivially on $A'=\mathrm{Hom}(A,\mu_p)$ since it acts trivially on A but non-trivially on μ_p. Therefore A'^{Δ} is a proper $\mathcal{G}$-submodule of A' so that $A'^{\Delta}=0$. This shows that $H^i(\mathcal{G},A')=0$ for all i.

The five term exact sequence belonging to the exact sequence $1\to G_{\mathcal{K}}\to G_k\to\mathcal{G}\to 1$ now yields the isomorphism

$$H^1(k,A')\xrightarrow{\sim} H^1(\mathcal{K},A')^{\mathcal{G}},$$

and the commutative diagram

$$\begin{array}{ccc}
H^1(k,A') & \longrightarrow & \prod\limits_{\mathfrak{p}\in P} H^1(k_{\mathfrak{p}},A') \\
\big\downarrow{\scriptstyle\cong} & & \big\downarrow \\
H^1(\mathcal{K},A')^{\mathcal{G}} & \longrightarrow & \prod\limits_{\mathfrak{P}|\mathfrak{p}\in P} H^1(\mathcal{K}_{\mathfrak{P}},A')
\end{array}$$

shows that it suffices to prove the injectivity of the lower map. Let $x\in H^1(\mathcal{K},A')^{\mathcal{G}}$ be in the kernel of this map. Since A' is a trivial $G_{\mathcal{K}}$-module, x is a $\mathcal{G}$-homomorphism $x\colon G_{\mathcal{K}}\to A'$ and defines an extension $\mathcal{N}|\mathcal{K}$ which is Galois over k. We have $x_{\mathfrak{P}}=0$ and thus $\mathcal{N}_{\mathfrak{P}}=\mathcal{K}_{\mathfrak{P}}$ for all $\mathfrak{P}|\mathfrak{p}\in P$. Let $P(\mathcal{K}|k)$ resp. $P(\mathcal{N}|k)$ be the set of primes which split completely in $\mathcal{K}$ resp. in $\mathcal{N}$. Since P contains $P(\mathcal{K}|k)$ up to a finite number of primes we see that almost all of the primes in $P(\mathcal{K}|k)$ are contained in $P(\mathcal{N}|k)$. This implies that

$$\frac{1}{[\mathcal{K}:k]}=\text{density of } P(\mathcal{K}|k)\leqq\text{density of } P(\mathcal{N}|k)=\frac{1}{[\mathcal{N}:k]},$$

i.e. $\mathcal{N}=\mathcal{K}$ and therefore $x=0$. This shows the required injectivity.

Lemma 3. *Let A be a simple G_k-module and $K|k$ a Galois extension such that $k(A)\subseteq K$ but $\mu_p\nsubseteq K$.*

Let S be a finite set of primes of k and $y_{\mathfrak{p}}\in H^1(k_{\mathfrak{p}},A)$ for $\mathfrak{p}\in S$. Then there exists an element $z\in H^1(k,A)$ with the properties

(i) $z_{\mathfrak{p}} = y_{\mathfrak{p}}$ *for* $\mathfrak{p} \in S$.

(ii) *If* $z_{\mathfrak{p}}$ *is ramified for* $\mathfrak{p} \notin S$ *then* $\mathfrak{p}$ *splits completely in* $K(\mu_p)$.

Proof. We can assume that S contains all primes above p and all infinite primes. Let $P = P(\mathscr{K}|k) \cup S$, where $P(\mathscr{K}|k)$ is the set of primes of k which split completely in $\mathscr{K} = K(\mu_p)$. It is clear that the lemma follows from the surjectivity of the map

$$(9) \quad H^1(k, A) \to \prod_{\mathfrak{p} \in S} H^1(k_{\mathfrak{p}}, A) \times \prod_{\mathfrak{p} \notin P} H^1(k_{\mathfrak{p}}, A)/H^1_{nr}(k_{\mathfrak{p}}, A).$$

In order to prove this surjectivity we use Lemma 2 and the duality theorem of Tate and Poitou. Let

$$X = \prod_{\mathfrak{p}} H^1(k_{\mathfrak{p}}, A) \quad \text{and} \quad X' = \prod_{\mathfrak{p}} H^1(k_{\mathfrak{p}}, A').$$

By the local duality theorem we have a duality

$$X \times X' \to \mathbb{Z}/p$$

of locally compact abelian groups. Let Y, Y' be the image of $H^1(k, A)$, $H^1(k, A')$ in X, X' respectively. By the global duality theorem Y' is the orthogonal complement of Y: $Y' = Y^{\perp}$.

Let us consider the following subgroup of X,

$$\Lambda = \prod_{\mathfrak{p} \in S} \{0_{\mathfrak{p}}\} \times \prod_{\mathfrak{p} \in P - S} H^1(k_{\mathfrak{p}}, A) \times \prod_{\mathfrak{p} \notin P} H^1_{nr}(k_{\mathfrak{p}}, A).$$

Then X/Λ is the group which stands on the right hand of the map (9), and therefore the surjectivity of this map is equivalent with

$$X = Y + \Lambda.$$

In order to prove this equality we first remark that $Y + \Lambda$ is a closed subgroup of X. Namely, the open subgroups of X are of the form

$$\prod_{\mathfrak{p} \in T} V_{\mathfrak{p}} \times \prod_{\mathfrak{p} \notin T} H^1_{nr}(k_{\mathfrak{p}}, A),$$

where T is a finite set of primes and $V_{\mathfrak{p}}$ a subgroup of $H^1(k_{\mathfrak{p}}, A)$. Since Λ and therefore also $Y + \Lambda$ contains such an open subgroup, $Y + \Lambda$ is an open and therefore closed subgroup of X. It follows that $X = Y + \Lambda$ is equivalent with

$$(Y + \Lambda)^{\perp} = Y^{\perp} \cap \Lambda^{\perp} = Y' \cap \Lambda^{\perp} = 0,$$

or in other words equivalent with the injectivity of the map $Y' \to X'/\Lambda^{\perp}$. Now we have

$$\Lambda^{\perp} = \prod_{\mathfrak{p} \in S} H^1(k_{\mathfrak{p}}, A') \times \prod_{\mathfrak{p} \in P - S} \{0_{\mathfrak{p}}\} \times \prod_{\mathfrak{p} \notin P} H^1_{nr}(k_{\mathfrak{p}}, A'),$$

and it is therefore sufficient to show the injectivity of

$$H^1(k, A') \to \prod_{\mathfrak{p} \in P - S} H^1(k_{\mathfrak{p}}, A') \times \prod_{\mathfrak{p} \notin P} H^1(k_{\mathfrak{p}}, A')/H^1_{nr}(k_{\mathfrak{p}}, A').$$

This however is clear since Lemma 2 asserts already the injectivity of the map

$$H^1(k, A') \to \prod_{\mathfrak{p} \in P-S} H^1(k_\mathfrak{p}, A').$$

Theorem 1. *Let A be a simple G_k-module with $pA=0$. Let $K|k$ be a Galois extension such that $k(A) \subseteq K$ but $\mu_p \nsubseteq K$ and let $\Omega|K$ be an abelian extension.*

Let S be a finite set of primes of k and $y_\mathfrak{p} \in H^1(k_\mathfrak{p}, A)$ for $\mathfrak{p} \in S$. Then there exists an element $x \in H^1(k, A)$ such that

(i) $x_\mathfrak{p} = y_\mathfrak{p}$ *for* $\mathfrak{p} \in S$.

(ii) *If* $\mathfrak{p} \notin S$, *then* $x_\mathfrak{p}$ *is cyclic and if* $x_\mathfrak{p}$ *is ramified then* $\mathfrak{p}$ *splits completely in* Ω.

Proof. Consider the commutative diagram

$$\begin{array}{ccc} H^1(K, A) & \longrightarrow & \prod\limits_{\mathfrak{p}} \prod\limits_{\mathfrak{P}|\mathfrak{p}} H^1(K_\mathfrak{P}, A) \\ \Big\downarrow \scriptstyle{\mathrm{cor}} & & \Big\downarrow \scriptstyle{\prod\limits_{\mathfrak{p}} \sum\limits_{\mathfrak{P}|\mathfrak{p}} \mathrm{cor}_\mathfrak{P}} \\ H^1(k, A) & \longrightarrow & \prod\limits_{\mathfrak{p}} H^1(k_\mathfrak{p}, A) \end{array}$$

where cor and $\mathrm{cor}_\mathfrak{P}$ are the corestriction maps. We want to apply the main lemma to the upper homomorphism. For this reason we would like to lift the elements $y_\mathfrak{p}$, $\mathfrak{p} \in S$, to elements of $\prod\limits_{\mathfrak{P}|\mathfrak{p}} H^1(K_\mathfrak{P}, A)$. However the map $\sum\limits_{\mathfrak{P}|\mathfrak{p}} \mathrm{cor}_\mathfrak{P}$ is in general not surjective. Therefore we first shift the local prescriptions $y_\mathfrak{p}$, $\mathfrak{p} \in S$, to other places, where they can be lifted.

By Lemma 3, there exists an element $z \in H^1(k, A)$ such that

(10) $z_\mathfrak{p} = y_\mathfrak{p}$ for $\mathfrak{p} \in S$,

(11) If $z_\mathfrak{p}$ is ramified for $\mathfrak{p} \notin S$ then $\mathfrak{p}$ splits completely in K.

Let $V = S \cup \{\mathfrak{p} \mid z_\mathfrak{p}$ is ramified$\}$ and define the elements $\eta_\mathfrak{p} \in H^1(k_\mathfrak{p}, A)$ for $\mathfrak{p} \in V$ by

$$\eta_\mathfrak{p} = 0 \quad \text{for } \mathfrak{p} \in S, \qquad \eta_\mathfrak{p} = -z_\mathfrak{p} \quad \text{for } \mathfrak{p} \in V - S.$$

Now these elements lie in the image of

$$\mathrm{cor}_\mathfrak{p} = \sum_{\mathfrak{P}|\mathfrak{p}} \mathrm{cor}_\mathfrak{P} \colon \prod_{\mathfrak{P}|\mathfrak{p}} H^1(K_\mathfrak{P}, A) \to H^1(k_\mathfrak{p}, A).$$

For $\mathfrak{p} \in S$ this is clear since $\eta_\mathfrak{p} = 0$ and for $\mathfrak{p} \in V - S$ we have $K_\mathfrak{p} = k_\mathfrak{p}$ so that $\mathrm{cor}_\mathfrak{p}$ is the usual sum

$$\sum_{\mathfrak{P}|\mathfrak{p}} \colon \prod_{\mathfrak{P}|\mathfrak{p}} H^1(k_\mathfrak{p}, A) \to H^1(k_\mathfrak{p}, A).$$

We now choose a pre-image $\tilde{\eta}_\mathfrak{p} = (\tilde{\eta}_\mathfrak{P})_{\mathfrak{P}|\mathfrak{p}} \in \prod\limits_{\mathfrak{P}|\mathfrak{p}} H^1(K_\mathfrak{P}, A)$ of $\eta_\mathfrak{p} \in H^1(k_\mathfrak{p}, A)$ for each $\mathfrak{p} \in V$. Before we apply the main lemma to these elements we enlarge the abelian extension $\Omega|K$ in the following way. Let z' be the image of $z \in H^1(k, A)$ under the restriction map

$$H^1(k, A) \to H^1(K, A).$$

Then z' is a homomorphism $z'\colon G_K \to A$ and defines an abelian extension $K'|K$. We set $\Omega' = K'\Omega$.

Now by the main lemma there exists an element $\tilde{\xi} \in H^1(K, A)$ such that

(12) $\tilde{\xi}_{\mathfrak{P}} = \tilde{\eta}_{\mathfrak{P}}$ for $\mathfrak{P}|\mathfrak{p} \in V$,

(13) If $\mathfrak{P}|\mathfrak{p} \notin V$ then $\tilde{\xi}_{\mathfrak{P}}$ is cyclic and if $\tilde{\xi}_{\mathfrak{P}}$ is ramified then $\mathfrak{p}$ splits completely in Ω' and $\tilde{\xi}_{\mathfrak{P}'} = 0$ for $\mathfrak{P}'|\mathfrak{p}$, $\mathfrak{P}' \neq \mathfrak{P}$.

Let ξ resp. $\xi_{\mathfrak{p}}$ the image of $\tilde{\xi}$ resp. $\tilde{\xi}_{\mathfrak{p}} = (\tilde{\xi}_{\mathfrak{P}})_{\mathfrak{P}|\mathfrak{p}}$ in $H^1(k, A)$ resp. $H^1(k_{\mathfrak{p}}, A)$. We show that element

$$x = \xi + z \in H^1(k, A)$$

satisfies the conditions of the theorem.

If $\mathfrak{p} \in S$ then

$$x_{\mathfrak{p}} = \xi_{\mathfrak{p}} + z_{\mathfrak{p}} = \eta_{\mathfrak{p}} + y_{\mathfrak{p}} = y_{\mathfrak{p}},$$

so that condition (i) is satisfied. If $\mathfrak{p} \in V - S$, then

$$x_{\mathfrak{p}} = \xi_{\mathfrak{p}} + z_{\mathfrak{p}} = \eta_{\mathfrak{p}} + z_{\mathfrak{p}} = -z_{\mathfrak{p}} + z_{\mathfrak{p}} = 0,$$

so that $x_{\mathfrak{p}}$ is unramified and cyclic. Let $\mathfrak{p} \notin V$. If $x_{\mathfrak{p}}$ is unramified then it is cyclic, since it becomes zero over an unramified extension of $k_{\mathfrak{p}}$, which is cyclic. So assume that $x_{\mathfrak{p}}$ is ramified. Since $\mathfrak{p} \notin V$, $z_{\mathfrak{p}}$ is unramified. Because $x_{\mathfrak{p}} = \xi_{\mathfrak{p}} + z_{\mathfrak{p}}$, $\xi_{\mathfrak{p}}$ must be ramified. This means that $\tilde{\xi}_{\mathfrak{P}}$ is ramified for one prime $\mathfrak{P}|\mathfrak{p}$ of K. Therefore by (13) $\mathfrak{p}$ splits completely in Ω'. In particular $\mathfrak{P}$ splits completely in K' and this means that $z_{\mathfrak{p}} = 0$, since, by definition of K', z becomes zero in $H^1(K', A)$ and therefore $z_{\mathfrak{p}}$ is zero in $H^1(K'_{\mathfrak{P}'}, A) = H^1(k_{\mathfrak{p}}, A)$, $\mathfrak{P}'$ being a prime of K' above $\mathfrak{p}$. We thus have $x_{\mathfrak{p}} = \xi_{\mathfrak{p}}$, i.e. $x_{\mathfrak{p}}$ is the image of $(\tilde{\xi}_{\mathfrak{P}})_{\mathfrak{P}|\mathfrak{p}}$ under the map

$$\sum_{\mathfrak{P}|\mathfrak{p}} \mathrm{cor}_{\mathfrak{P}}\colon \prod_{\mathfrak{P}|\mathfrak{p}} H^1(K_{\mathfrak{P}}, A) \to H^1(k_{\mathfrak{p}}, A).$$

Since $\mathfrak{p}$ splits completely in K we have $K_{\mathfrak{P}} = k_{\mathfrak{p}}$ and this map is the sum

$$\sum_{\mathfrak{P}|\mathfrak{p}}\colon \prod_{\mathfrak{P}|\mathfrak{p}} H^1(K_{\mathfrak{P}}, A) \to H^1(k_{\mathfrak{p}}, A),$$

i.e. $x_{\mathfrak{p}} = \sum_{\mathfrak{P}|\mathfrak{p}} \tilde{\xi}_{\mathfrak{P}} = \tilde{\xi}_{\mathfrak{P}}$ since $\tilde{\xi}_{\mathfrak{P}'} = 0$ for $\mathfrak{P}'|\mathfrak{p}$, $\mathfrak{P}' \neq \mathfrak{P}$. Since $\tilde{\xi}_{\mathfrak{P}}$ is cyclic so is $x_{\mathfrak{p}}$. This proves the theorem.

We conclude this paragraph with the following local-global-principle, which was pointed out to us by O. Neumann (see also [10]).

Theorem 2. *Let A be a simple G_k-module such that $pA = 0$ and $\mu_p \nsubseteq k(A)$. Then the homomorphism*

$$H^2(k, A) \to \prod_{\mathfrak{p}} H^2(k_{\mathfrak{p}}, A)$$

is injective.

Proof. By Lemma 2 the homomorphism $H^1(k, A') \to \prod\limits_{\mathfrak{p}} H^1(k_{\mathfrak{p}}, A')$ is injective and therefore the theorem follows from the duality theorem of Tate and Poitou.

§ 2. The Main Theorem and its Proof

In this paragraph we prove a general theorem about the absolute Galois group of a number field and we shall explain its field theoretical meaning and its consequences in the next paragraph. We shall suppose in the sequel that all occuring profinite groups are separable, i.e. have a countable basis of neighbourhoods of the identity. We say that a profinite group is of finite exponent, if its elements have a finite bounded order. The exponent of G is then the l.c.m. of these orders.

Let Γ be a fixed finite group. The homomorphisms $G \xrightarrow{f} \Gamma$ of arbitrary profinite groups into Γ are the objects of category, if one defines as morphisms from $G \xrightarrow{f} \Gamma$ to $G' \xrightarrow{f'} \Gamma$ all homomorphisms $G \xrightarrow{\psi} G'$ with $f' \circ \psi = f$. We call two such morphisms

$$\begin{tikzcd} G \arrow[rr, shift left, "\psi"] \arrow[rr, shift right, "\psi'"'] \arrow[dr, "f"'] & & G' \arrow[dl, "f'"] \\ & \Gamma & \end{tikzcd}$$

equivalent, $\psi \sim \psi'$, if there exists a fixed element $a \in \ker(f')$ such that

$$\psi'(\sigma) = a\psi(\sigma)a^{-1} \qquad \text{for all } \sigma \in G,$$

i.e. if ψ and ψ' are conjugate under an element of $\ker(f')$. We denote by $[\psi]$ the equivalence classes and by $\mathcal{H}om_\Gamma(G, G')$ the set of all equivalence classes $[\psi]$. Furthermore we denote by $\mathcal{H}om_\Gamma(G, G')_{\mathrm{sur}}$ the subset of all $[\psi]$ with surjective $\psi: G \to G'$.

Let now k be an algebraic number field and $k_{\mathfrak{p}}$ its completions at the primes $\mathfrak{p}$ of k. By $\mathfrak{G}$ resp. $\mathfrak{G}_{\mathfrak{p}}$ we denote the Galois group of an algebraic closure $\bar{k}$ of k resp. $\bar{k}_{\mathfrak{p}}$ of $k_{\mathfrak{p}}$. For each prime $\mathfrak{p}$ we choose a fixed k-imbedding $\bar{k} \to \bar{k}_{\mathfrak{p}}$ and obtain in this way imbeddings $\mathfrak{G}_{\mathfrak{p}} \hookrightarrow \mathfrak{G}$.

Let $\mathfrak{G} \xrightarrow{\varphi} \Gamma$ be a homomorphism of $\mathfrak{G}$ into the finite group Γ and $\mathfrak{G}_{\mathfrak{p}} \xrightarrow{\varphi_{\mathfrak{p}}} \Gamma$ its restriction to $\mathfrak{G}_{\mathfrak{p}}$. If then $G \xrightarrow{f} \Gamma$ is a homomorphism of an arbitrary profinite group G, we obtain the diagrams

$$\begin{tikzcd} \mathfrak{G} \arrow[rr, dashed, "\psi"] \arrow[dr, "\varphi"'] & & G \arrow[dl, "f"] \\ & \Gamma & \end{tikzcd} \qquad \begin{tikzcd} \mathfrak{G}_{\mathfrak{p}} \arrow[rr, dashed, "\psi_{\mathfrak{p}}"] \arrow[dr, "\varphi_{\mathfrak{p}}"'] & & G \arrow[dl, "f"] \\ & \Gamma & \end{tikzcd}$$

and the canonical restriction map

$$\mathcal{H}om_\Gamma(\mathfrak{G}, G) \to \prod_{\mathfrak{p}} \mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G).$$

Main Theorem. *Let $\mathfrak{G} \xrightarrow{\varphi} \Gamma$ be a surjective homomorphism onto the finite group Γ, and let $m(K)$ be the number of roots of unity in the fixed field K of the kernel of φ.*

If $G \xrightarrow{f} \Gamma$ is a surjective homomorphism with a pro-solvable kernel of finite exponent, which is prime to $m(K)$, and if

$$\prod_{\mathfrak{p}} \mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G) \neq \emptyset,$$

then the map

$$\mathcal{H}om_\Gamma(\mathfrak{G}, G)_{\mathrm{sur}} \to \prod_{\mathfrak{p} \in S} \mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)$$

is surjective for every finite set S of primes of k.

Remark. It is easy to see that the condition $\mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G) \neq \emptyset$ is automatically satisfied for almost all primes, namely for all $\mathfrak{p}$ which are unramified in K (see Lemma 5).

We shall deduce the proof of the theorem from the more special case, where the kernel of the homomorphism $G \xrightarrow{f} \Gamma$ is a finite abelian group A. Therefore we study a fixed group extension

$$1 \to A \to G \to \Gamma \to 1$$

and a surjective homomorphism $\varphi: \mathfrak{G} \to \Gamma$. Then Γ acts on A by inner automorphisms and A becomes a $\mathfrak{G}$-module via the homomorphism φ.

The cohomology group $H^1(\mathfrak{G}, A)$ acts on the set $\mathcal{H}om_\Gamma(\mathfrak{G}, G)$ in the following way. Let $[\psi] \in \mathcal{H}om_\Gamma(\mathfrak{G}, G)$ and $x \in H^1(\mathfrak{G}, A)$. Let $\psi: \mathfrak{G} \to G$ be a representative of $[\psi]$ and $\chi: \mathfrak{G} \to A$ a 1-cocycle in the cohomology class x. Then the function

$$\psi'(\sigma) = \chi(\sigma) \cdot \psi(\sigma), \qquad \sigma \in \mathfrak{G},$$

is again a homomorphism $\psi': \mathfrak{G} \to G$ such that $f \circ \psi' = \varphi$. If $\chi(\sigma) = a^\sigma a^{-1}$, $a \in A$, is a coboundary then $a^\sigma = \psi(\sigma)\, a \psi(\sigma)^{-1}$ and

$$\psi'(\sigma) = a^{-1} \psi(\sigma)\, a \psi(\sigma)^{-1}\, \psi(\sigma) = a^{-1} \psi(\sigma)\, a,$$

i.e. $\psi' \sim \psi$. Therefore the class $[\psi]^x = [\psi'] \in \mathcal{H}om_\Gamma(\mathfrak{G}, G)$ is independent of the choice of $\psi \in [\psi]$ and $\chi \in x$.

One verifies immediately that $\mathcal{H}om_\Gamma(\mathfrak{G}, G)$ becomes a *principal homogeneous space* over the group $H^1(\mathfrak{G}, A)$. Similarly $\mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, A)$ is a principal homogeneous space over $H^1(\mathfrak{G}_\mathfrak{p}, A)$.

An element $[\psi] \in \mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)$ is called unramified if the inertia group $\mathfrak{T}_\mathfrak{p}$ of $\mathfrak{G}_\mathfrak{p}$ lies in the kernel of ψ. Here we set $\mathfrak{T}_\mathfrak{p} = \mathfrak{G}_\mathfrak{p}$ if $\mathfrak{p}$ is an infinite prime. By

$\mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)_{nr}$ we denote the set of unramified elements of $\mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)$, and by $\prod_\mathfrak{p} \mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)$ the restricted product with respect to these subsets. Since the sets $\mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)$ are finite, this product is a locally compact topological space. We have the

Proposition. *The pair of maps*

$$\mathcal{H}om_\Gamma(\mathfrak{G}, G) \to \prod_\mathfrak{p} \mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G), \qquad H^1(\mathfrak{G}, A) \to \prod_\mathfrak{p} H^1(\mathfrak{G}_\mathfrak{p}, A)$$

is a morphism of locally compact principal homogeneous spaces.

For the following considerations we assume that A is a *simple* $\mathfrak{G}$-module $\neq 0$. Then A is annihilated by a prime number p, $pA=0$, and the condition in the theorem says that the field K, which is defined by the homomorphism $\varphi\colon \mathfrak{G} \to \Gamma$ does not contain the group μ_p of p-th roots of unity. We next establish the following local-global-principle.

Lemma 4. *Let the kernel A of $f\colon G \to \Gamma$ be a simple Γ-module such that $pA=0$ and $\mu_p \nsubseteq K$. Then*

$$\mathcal{H}om_\Gamma(\mathfrak{G}, G) \neq \emptyset \quad \Leftrightarrow \quad \prod_\mathfrak{p} \mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G) \neq \emptyset.$$

Proof. If $\hat{\mathfrak{G}} = \mathfrak{G} \times_\Gamma G = \{(\sigma, \tau) \in \mathfrak{G} \times \Gamma \mid \varphi(\sigma) = f(\tau)\}$ is the fiber product of $\mathfrak{G}$ and G over Γ, then we have the commutative diagram

$$\begin{array}{ccccccccc}
1 & \longrightarrow & A & \longrightarrow & \hat{\mathfrak{G}} & \xrightarrow{\hat{f}} & \mathfrak{G} & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow{\scriptstyle\varphi} & & \\
1 & \longrightarrow & A & \longrightarrow & G & \xrightarrow{f} & \Gamma & \longrightarrow & 1.
\end{array}$$

Because of the universal property of the fiber product the homomorphisms $\psi\colon \mathfrak{G} \to G$ with $f \circ \psi = \varphi$ are in 1-1-correspondence with the homomorphisms $s\colon \mathfrak{G} \to \hat{\mathfrak{G}}$ with $\hat{f} \circ s = \mathrm{id}_\mathfrak{G}$. Therefore $\mathcal{H}om_\Gamma(\mathfrak{G}, G) \neq \emptyset$ if and only if the upper group extension splits. Let us consider now the homomorphisms

$$\varphi^*\colon H^2(\Gamma, A) \to H^2(\mathfrak{G}, A) \quad \text{resp.} \quad \varphi_\mathfrak{p}^*\colon H^2(\Gamma, A) \to H^2(\mathfrak{G}_\mathfrak{p}, A),$$

which are induced by $\varphi\colon \mathfrak{G} \to \Gamma$ resp. $\varphi_\mathfrak{p}\colon \mathfrak{G}_\mathfrak{p} \to \Gamma$. If $x \in H^2(\Gamma, A)$ is the class belonging to the lower group extension then $\varphi^*(x) \in H^2(\mathfrak{G}, A)$ belongs to the upper group extension. It thus follows that $\mathcal{H}om_\Gamma(\mathfrak{G}, G) \neq \emptyset \Leftrightarrow \varphi^*(x) = 0$, and in the same way $\mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G) \neq \emptyset \Leftrightarrow \varphi_\mathfrak{p}^*(x) = 0$. If we now consider the commutative diagram

$$\begin{array}{ccc}
 & H^2(\Gamma, A) & \\
{\scriptstyle\varphi^*}\swarrow & & \searrow{\scriptstyle\prod \varphi_\mathfrak{p}^*} \\
H^2(\mathfrak{G}, A) & \longrightarrow & \prod_\mathfrak{p} H^2(\mathfrak{G}_\mathfrak{p}, A)
\end{array}$$

and regard that by Theorem 2 of § 1 the horizontal arrow is injective, we see that

$$\prod_{\mathfrak{p}} \mathcal{H}om_{\Gamma}(\mathfrak{G}_{\mathfrak{p}}, G) \neq \emptyset \Leftrightarrow \prod_{\mathfrak{p}} \varphi_{\mathfrak{p}}^*(x) = 0 \Leftrightarrow \varphi^*(x) = 0 \Leftrightarrow \mathcal{H}om_{\Gamma}(\mathfrak{G}, G) \neq \emptyset.$$

Lemma 5. *Let $G_2 \to G_1$ be a surjective homomorphism between arbitrary profinite groups and let $\mathfrak{G}_{\mathfrak{p}} \to G_1$ be an unramified homomorphism.*

Then $\mathcal{H}om_{G_1}(\mathfrak{G}_{\mathfrak{p}}, G_2)_{nr}$ and a fortiori $\mathcal{H}om_{G_1}(\mathfrak{G}_{\mathfrak{p}}, G_2)$ is not empty.

Proof. If the homomorphism $\mathfrak{G}_{\mathfrak{p}} \to G_1$ is unramified, then it factorizes over the factor group $\mathfrak{G}_{\mathfrak{p}}/\mathfrak{T}_{\mathfrak{p}}$. But this factor group is isomorphic to the free profinite group $\hat{\mathbb{Z}}$ of rank 1. Therefore the homomorphism $\mathfrak{G}_{\mathfrak{p}}/\mathfrak{T}_{\mathfrak{p}} \to G_1$ can be lifted to a homomorphism $\mathfrak{G}_{\mathfrak{p}}/\mathfrak{T}_{\mathfrak{p}} \to G_2$, i.e. the homomorphism $\mathfrak{G}_{\mathfrak{p}} \to G_1$ can be lifted to an unramified homomorphism $\mathfrak{G}_{\mathfrak{p}} \to G_2$.

We now come to the essential step in the proof of the main theorem. We consider the case that the kernel of the group extension $f: G \to \Gamma$ is a simple Γ-module A. By the local-global-principle of Lemma 4 there exists a Γ-homomorphism $\psi: \mathfrak{G} \to G$, since we assume $\mathcal{H}om_{\Gamma}(\mathfrak{G}_{\mathfrak{p}}, G) \neq \emptyset$ for all $\mathfrak{p}$. Suppose for a moment that we can find ψ (surjective) in such a way that it induces given Γ-homomorphisms $\psi_{\mathfrak{p}}: \mathfrak{G}_{\mathfrak{p}} \to G$ at the primes $\mathfrak{p}$ in the finite set S. Clearly then the theorem is true for this case. However this is not enough and we need a much sharper result. Namely having found the homomorphism $\psi: \mathfrak{G} \to G$, we shall have to lift it in a next step to a larger group extension $G' \to G \to 1$. In view of the local-global principle we will therefore have to impose on ψ the condition, that *all* its local restrictions $\psi_{\mathfrak{p}}: \mathfrak{G}_{\mathfrak{p}} \to G$ can be lifted to G'. For a fixed finite set of primes (not depending on ψ) this causes no essential difficulties. The essential problem occurs in the following way. Let $N|k$ be the Galois extension defined by the homomorphism $\psi: \mathfrak{G} \to G$. Then the completion $N_{\mathfrak{p}}|k_{\mathfrak{p}}$ is the extension defined by the restriction $\psi_{\mathfrak{p}}: \mathfrak{G}_{\mathfrak{p}} \to G$. The lifting problem for $\psi_{\mathfrak{p}}$ then means, that the field $N_{\mathfrak{p}}|k_{\mathfrak{p}}$ can be imbedded into a larger extension with a given Galois group which is defined by the group extension $G' \to G$. If $N_{\mathfrak{p}}|k_{\mathfrak{p}}$ is unramified this can be always done. If however $N_{\mathfrak{p}}|k_{\mathfrak{p}}$ is ramified, i.e. if $\psi_{\mathfrak{p}}$ is ramified, then this is in general possible only under the conditions

1) $N_{\mathfrak{p}}|k_{\mathfrak{p}}$ is cyclic,

2) $k_{\mathfrak{p}}$ contains enough roots of unity, i.e. $\mathfrak{p}$ splits completely in a field $\Omega = k(\zeta_n)$ of roots of unity.

We therefore have to find the homomorphism ψ in such a way that its local restrictions $\psi_{\mathfrak{p}}$ satisfy these conditions at the primes $\mathfrak{p}$, where $\psi_{\mathfrak{p}}$ ramifies. This leads to the study of the following situation.

Besides of the group extension

$$1 \to A \to G \to \Gamma \to 1$$

with the simple Γ-module A $(pA = 0)$, and the homomorphism $\psi: \mathfrak{G} \to \Gamma$ we consider a further surjective homomorphism

$$E \xrightarrow{g} G$$

with a pro-solvable kernel of finite exponent e. Let n be a multiple of ep. We then prove,

Lemma 6. *Let S be an arbitrary finite set of primes of k and assume that $(n, m(K))=1$. If $\prod_{\mathfrak{p}} \mathscr{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G) \neq \emptyset$, then there exists an element $[\psi] \in \mathscr{H}om_\Gamma(\mathfrak{G}, G)_{\mathrm{sur}}$ with the following properties:*

(i) *$[\psi]$ induces given elements $[\psi_\mathfrak{p}] \in \mathscr{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)$ at the primes $\mathfrak{p} \in S$.*

(ii) *If $\mathfrak{p} \notin S$ is unramified in $K|k$, then $\mathscr{H}om_G(\mathfrak{G}_\mathfrak{p}, E) \neq \emptyset$,* i.e. *$\psi_\mathfrak{p}$ can be lifted to a G-homomorphism $\mathfrak{G}_\mathfrak{p} \to E$.*

(iii) *For the field N defined by $\psi: \mathfrak{G} \to G$ we have $(n, m(N))=1$.*

Proof. The proof will be in four steps, where the first three steps are reductions.

Step 1. We first show that we can assume that S contains all ramification primes of $K|k$, all primes dividing n and all infinite primes. Let S^* be the union of S with all these primes. Then for every $\mathfrak{p} \in S^* \setminus S$ we can add an element $[\psi_\mathfrak{p}]$ from the non-empty set $\mathscr{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)$ to the $[\psi_\mathfrak{p}]$, which are already given. If $\mathfrak{p} \in S^* \setminus S$ is unramified in $K|k$, then by Lemma 5 we can choose $[\psi_\mathfrak{p}]$ to be unramified. If now $[\psi] \in \mathscr{H}om_\Gamma(\mathfrak{G}, G)_{\mathrm{sur}}$ satisfies conditions (i)–(iii) for S^* instead of S, then so it does for the set S. In fact, (i) and (iii) are trivially satisfied, and since for the unramified primes $\mathfrak{p} \in S^* \setminus S$ we have chosen unramified elements $[\psi_\mathfrak{p}]$, condition (ii) follows from Lemma 5.

Step 2. We next show that it suffices to prove the existence of an element $[\psi] \in \mathscr{H}om_\Gamma(\mathfrak{G}, G)$, i.e. that we can drop the surjectivity of ψ. Let $\mathfrak{q} \notin S$ be a prime which splits completely in $K|k$, and let $a \in A$, $a \neq 0$. Let $\psi_\mathfrak{q}: \mathfrak{G}_\mathfrak{q} \to (a)$ be an unramified surjective homomorphism onto the cyclic subgroup $(a) \subseteq G$ generated by a. Since $\mathfrak{q}$ splits completely in K we have $\varphi_\mathfrak{q}(\mathfrak{G}_\mathfrak{q}) = \{1\}$ and hence $[\psi_\mathfrak{q}] \in \mathscr{H}om_\Gamma(\mathfrak{G}_\mathfrak{q}, G)$. If we now replace S by $S^* = S \cup \{\mathfrak{q}\}$ and the system $([\psi_\mathfrak{p}])_{\mathfrak{p} \in S}$ by the system $([\psi_\mathfrak{p}])_{\mathfrak{p} \in S^*}$, then an element $[\psi] \in \mathscr{H}om_\Gamma(\mathfrak{G}, G)$, which satisfies the conditions (i)–(iii) for S^* instead of S, satisfies them also for S, as is seen in the same way as in the first step. Moreover ψ is surjective. Namely, $\psi(\mathfrak{G}_\mathfrak{q}) \subseteq G$ is conjugate to $\psi_\mathfrak{q}(\mathfrak{G}_\mathfrak{q}) = (a)$ under an element of A, i.e. $a \in \psi(\mathfrak{G}_\mathfrak{q}) \subseteq \psi(\mathfrak{G})$. Therefore $A \cap \psi(\mathfrak{G})$ is a non-trivial Γ-submodule of A. Since A is simple this means that $A \subseteq \psi(\mathfrak{G})$ and hence $\psi(\mathfrak{G}) = G$.

Step 3. It suffices to prove the existence of an element $[\psi] \in \mathscr{H}om_\Gamma(\mathfrak{G}, G)$ satisfying only the conditions (i) and (ii). In order to see this, let ζ_n be a primitive n-th root of unity, and let $\sigma_1, \dots, \sigma_r$ be generators of $\mathrm{Gal}(K(\zeta_n)|K)$. Let $\mathfrak{p}_1, \dots, \mathfrak{p}_r \notin S$ be different primes such that $\mathfrak{p}_i$ splits completely in K and

$$\left(\frac{K(\zeta_n)|K}{\mathfrak{P}_i}\right) = \sigma_i, \qquad i = 1, \dots, r,$$

where $\mathfrak{P}_i$ is a prime of K above $\mathfrak{p}_i$. Then $\varphi_{\mathfrak{p}_i}$ is the trivial homomorphism, so that the trivial homomorphism $\psi_{\mathfrak{p}_i}: \mathfrak{G}_{\mathfrak{p}_i} \to G$ yields an element $[\psi_{\mathfrak{p}_i}] \in \mathscr{H}om_\Gamma(\mathfrak{G}_{\mathfrak{p}_i}, G)$, $i = 1, \dots, n$. Again we replace S by $S^* = S \cup \{\mathfrak{p}_1, \dots, \mathfrak{p}_r\}$ and $([\psi_\mathfrak{p}])_{\mathfrak{p} \in S}$ by $([\psi_\mathfrak{p}])_{\mathfrak{p} \in S^*}$. If now $[\psi] \in \mathscr{H}om_\Gamma(\mathfrak{G}, G)$ satisfies the conditions (i), (ii) for S^*, then so it does for S, as we have seen in the first reduction. Moreover it satisfies condition (iii). In fact, the restriction of ψ to $\mathfrak{G}_{\mathfrak{p}_i}$ is the trivial homomorphism $\psi_{\mathfrak{p}_i}$, and this means, that the primes $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ split completely in the field $N|k$, which is defined by the homomorphism $\psi\colon \mathfrak{G} \to G$. Therefore the automorphisms $\bar{\sigma}_i = \left(\frac{N \cap K(\zeta_n)|K}{\mathfrak{P}_i}\right)$, which generate the Galois group of $N \cap K(\zeta_n)|K$, are all one, so that $N \cap K(\zeta_n) = K$. Let now $d = (n, m(N))$ and let ζ_d be a primitive d-th root of unity. Then $\zeta_d \in N \cap K(\zeta_n) = K$, i.e. $d|(n, m(K))$ and hence $d = 1$.

Step 4. This is the essential step in the proof of the main theorem. By the Proposition of § 1 the maps

$$\mathscr{H}om_\Gamma(\mathfrak{G}, G) \to \prod_{\mathfrak{p}} \mathscr{H}om_\Gamma(\mathfrak{G}_{\mathfrak{p}}, G), \qquad H^1(\mathfrak{G}, A) \to \prod_{\mathfrak{p}} H^1(\mathfrak{G}_{\mathfrak{p}}, A)$$

form a morphism of principal homogeneous spaces. Since by assumption $\prod_{\mathfrak{p}} \mathscr{H}om_\Gamma(\mathfrak{G}_{\mathfrak{p}}, G) \neq \emptyset$ the set $\mathscr{H}om_\Gamma(\mathfrak{G}, G)$ is not empty by Lemma 4, and we can start with an element

$$[\psi_0] \in \mathscr{H}om_\Gamma(\mathfrak{G}, G).$$

Our aim is to change this element by a cohomology class $x \in H^1(\mathfrak{G}, A)$ into an element $[\psi]$ which satisfies the conditions of the lemma. Let $N_0|k$ be the field defined by the homomorphism $\psi_0\colon \mathfrak{G} \to G$ and let $\Omega = N_0(\zeta_n)$, where ζ_n is a primitive n-th root of unity. Then $\Omega|K$ is an abelian extension.

In order to avoid confusion we denote by $[\psi_{1\mathfrak{p}}] \in \mathscr{H}om_\Gamma(\mathfrak{G}_{\mathfrak{p}}, G)$ the elements which are given in advance for $\mathfrak{p} \in S$. Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ be the primes $\mathfrak{p} \notin S$ for which $\psi_{0\mathfrak{p}}$ is ramified. Then $\mathfrak{p}_i$ is unramified in $K|k$, i.e. the homomorphism $\varphi_{\mathfrak{p}_i}$ is unramified and can therefore be lifted to an unramified Γ-homomorphism $\psi_{1\mathfrak{p}_i}\colon \mathfrak{G}_{\mathfrak{p}_i} \to G$. Let $S^* = S \cup \{\mathfrak{p}_1, \ldots, \mathfrak{p}_r\}$. For each $\mathfrak{p} \in S^*$ let $y_{\mathfrak{p}} \in H^1(\mathfrak{G}_{\mathfrak{p}}, A)$ be the cohomology class which sends $[\psi_{0\mathfrak{p}}]$ into $[\psi_{1\mathfrak{p}}]$:

$$[\psi_{0\mathfrak{p}}]^{y_{\mathfrak{p}}} = [\psi_{1\mathfrak{p}}].$$

Since $p \nmid m(K)$ we have $\mu_p \nsubseteq K$. Therefore we can apply Theorem 1 of § 1 and find an element $x \in H^1(\mathfrak{G}, A)$ such that

(i) $x_{\mathfrak{p}} = y_{\mathfrak{p}}$ for $\mathfrak{p} \in S^*$,

(ii) If $\mathfrak{p} \notin S^*$ then $x_{\mathfrak{p}}$ is cyclic and if $x_{\mathfrak{p}}$ is ramified then $\mathfrak{p}$ splits completely in Ω.

We now prove that the element

$$[\psi] = [\psi_0]^x \in \mathscr{H}om_\Gamma(\mathfrak{G}, G)$$

satisfies the conditions (i) and (ii) of the lemma. For $\mathfrak{p} \in S^*$ we have

$$[\psi_{\mathfrak{p}}] = [\psi_{0\mathfrak{p}}]^{x_{\mathfrak{p}}} = [\psi_{0\mathfrak{p}}]^{y_{\mathfrak{p}}} = [\psi_{1\mathfrak{p}}].$$

Therefore condition (i) is satisfied. Let $\mathfrak{p} \notin S$ be unramified in $K|k$. If $\mathfrak{p} \in S^* \setminus S$ then $[\psi_{\mathfrak{p}}] = [\psi_{1\mathfrak{p}}]$ is unramified. Hence by Lemma 5, the homomorphism $\psi_{\mathfrak{p}}\colon \mathfrak{G}_{\mathfrak{p}} \to G$ can be lifted to a G-homomorphism $\mathfrak{G}_{\mathfrak{p}} \to E$, i.e. $\mathscr{H}om_G(\mathfrak{G}_{\mathfrak{p}}, E) \neq \emptyset$. Let $\mathfrak{p} \notin S^*$. If $[\psi_{\mathfrak{p}}]$ is unramified, then for the same reason $\mathscr{H}om_G(\mathfrak{G}\mathfrak{p}, E) \neq \emptyset$. Let us assume that $[\psi_\mathfrak{p}]=[\psi_{0\mathfrak{p}}]^{x_\mathfrak{p}}$ is ramified. Since $[\psi_{0\mathfrak{p}}]$ is unramified for $\mathfrak{p}\notin S^*$, the cohomology class $x_\mathfrak{p}$ must be ramified. Thus by (ii) of Theorem 1 $\mathfrak{p}$ splits completely in Ω. In particular, it splits completely in the field $N_0|k$, which is defined by ψ_0 and in $k(\zeta_n)$. This means that $\psi_{0\mathfrak{p}}\colon \mathfrak{G}_\mathfrak{p}\to G$ is the trivial homomorphism, or in other words that $N_{0\mathfrak{p}}=k_\mathfrak{p}$, and that $\zeta_n\in k_\mathfrak{p}$. Because of $K_\mathfrak{p}=k_\mathfrak{p}$ the Galois module A is a trivial $\mathfrak{G}_\mathfrak{p}$-module, i.e. $x_\mathfrak{p}$ is a homomorphism

$$x_\mathfrak{p}\colon \mathfrak{G}_\mathfrak{p}\to A\subseteq G,$$

namely it is the homomorphism, which represents the class $[\psi_{0\mathfrak{p}}]^{x_\mathfrak{p}}=[\psi_\mathfrak{p}]$. Therefore it remains only to show that $x_\mathfrak{p}$ can be lifted to a G-homomorphism $\mathfrak{G}_\mathfrak{p}\to E$. Since $x_\mathfrak{p}$ is cyclic it defines a cyclic extension $N_\mathfrak{p}\,|\,k_\mathfrak{p}$ which must be of degree p since $pA=0$. This extension is ramified and since $\mathfrak{p}\nmid p$ we have $N_\mathfrak{p}=k_\mathfrak{p}(\sqrt[p]{\pi})$ with a prime element π.

Let σ be an element of E which, under the map $E\to G$, is mapped onto a generator $\bar{\sigma}$ of

$$\mathrm{Gal}(N_\mathfrak{p}\,|\,k_\mathfrak{p})=x_\mathfrak{p}(\mathfrak{G}_\mathfrak{p})\subseteq G.$$

Since $\ker(E\to G)$ has exponent dividing n/p, the order m of σ is a divisor of n. Therefore $k_\mathfrak{p}$ contains the m-th roots of unity, so that $M_\mathfrak{p}=k_\mathfrak{p}(\sqrt[m]{\pi})\,|\,k_\mathfrak{p}$ is a normal and hence cyclic extension of degree m. Now consider the commutative diagram

$$\begin{tikzcd}
& \mathfrak{G}_\mathfrak{p} \arrow[dl] \arrow[d] \\
\mathrm{Gal}(M_\mathfrak{p}\,|\,k_\mathfrak{p}) \arrow[r] & \mathrm{Gal}(N_\mathfrak{p}\,|\,k_\mathfrak{p}) \arrow[d] \\
E \arrow[r, "g"] & G
\end{tikzcd}$$

in which the composite $\mathfrak{G}_\mathfrak{p}\to\mathrm{Gal}(N_\mathfrak{p}\,|\,k_\mathfrak{p})\to G$ is the homomorphism $x_\mathfrak{p}=\psi_\mathfrak{p}$. This diagram can be commutatively completed, by mapping a generator τ of $\mathrm{Gal}(M_\mathfrak{p}\,|\,k_\mathfrak{p})$ with image $\bar{\sigma}$ in $\mathrm{Gal}(N_\mathfrak{p}\,|\,k_\mathfrak{p})$ onto σ.

The proof of the lemma is now complete.

Proof of the Main Theorem. We consider the surjective homomorphism $\varphi\colon\mathfrak{G}\to\Gamma$ onto the finite group Γ and the group extension

$$1\to H\to G\xrightarrow{f}\Gamma\to 1,$$

where the kernel H is a pro-solvable group of finite exponent n and $(n, m(K))=1$. Since H is pro-solvable we find a chain

$$H=H_0\supseteq H_1\supseteq H_2\supseteq H_3\supseteq\dots$$

of open normal subgroups of H such that the factors $A_i=H_i/H_{i+1}$ are abelian and $\bigcap_i H_i=\{1\}$. Since Γ is finite we can assume that the H_i are normal in G.

Then the A_i are Γ-modules and become $\mathfrak{G}$-modules via φ. Taking a maximal refinement of such a chain we can assume that each A_i is a simple $\mathfrak{G}$-module.

For every $\mathfrak{p} \in S$ we prescribe an element $[\psi_\mathfrak{p}] \in \mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)$ and select a Γ-homomorphism $\psi_\mathfrak{p}\colon \mathfrak{G}_\mathfrak{p} \to G$ in $[\psi_\mathfrak{p}]$. We construct an element $[\psi] \in \mathcal{H}om_\Gamma(\mathfrak{G}, G)_{\mathrm{sur}}$ such that $[\psi | \mathfrak{G}_\mathfrak{p}] = [\psi_\mathfrak{p}]$ for $\mathfrak{p} \in S$. Let $G_i = G/H_i$ and consider the commutative diagrams

$$
\begin{tikzcd}
 & \mathfrak{G} \arrow[d, "\varphi"] \\
G \arrow[r, "f"] \arrow[d] & \Gamma \arrow[d] \\
G_i \arrow[r, "f_i"] \arrow[d, "\pi_i"'] & \Gamma \arrow[d] \\
G_{i-1} \arrow[r, "f_{i-1}"] & \Gamma
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \mathfrak{G}_\mathfrak{p} \arrow[dl, "\psi_\mathfrak{p}"'] \arrow[d, "\varphi_\mathfrak{p}"] \arrow[ddl, "\psi_{\mathfrak{p}i}"] \\
G \arrow[r] \arrow[d] & \Gamma \arrow[d] \\
G_i \arrow[r] \arrow[d, "\pi_i"'] & \Gamma \arrow[d] \\
G_{i-1} \arrow[r] & \Gamma,
\end{tikzcd}
\qquad \mathfrak{p} \in S.
$$

The kernel of $G_i \xrightarrow{f_i} \Gamma$ is H/H_i and the kernel of $G_i \xrightarrow{\pi_i} G_{i-1}$ is the simple $\mathfrak{G}$-module A_i. Let $p_i = \exp(A_i)$ and $n_i = \exp(H_i)$.

Claim. There exists a sequence of surjective Γ-homomorphisms $\psi_i\colon \mathfrak{G} \to G_i$ with the following properties

1) $\psi_{i-1} = \pi_i \circ \psi_i$, $i = 1, 2, \ldots$.
2) $[\psi_i | \mathfrak{G}_\mathfrak{p}] = [\psi_{\mathfrak{p}i}]$ in $\mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G_i)$ for $\mathfrak{p} \in S$.
3) For the field K_i defined by ψ_i we have $(n, m(K_i)) = 1$.
4) $\prod\limits_\mathfrak{p} \mathcal{H}om_{G_i}(\mathfrak{G}_\mathfrak{p}, G) \neq \emptyset$.

Proof of the Claim: If $i=0$, we can identify G_0 with Γ and choose $\psi_0 = \varphi$. The first condition is empty, the second is true since $\varphi|_{\mathfrak{G}_\mathfrak{p}} = \varphi_\mathfrak{p}$, and the third and fourth are an assumption of the theorem.

Let us assume, that the homomorphism $\psi_{i-1}\colon \mathfrak{G} \to G_{i-1}$ is already constructed. Let S_{i-1} be the union of S with all primes of k which ramify in the field $K_{i-1}|k$, defined by ψ_{i-1}. For each $\mathfrak{p} \in S_{i-1} \setminus S$ we can choose a G_{i-1}-homomorphism $\psi_\mathfrak{p}\colon \mathfrak{G}_\mathfrak{p} \to G$ since the set $\mathcal{H}om_{G_{i-1}}(\mathfrak{G}_\mathfrak{p}, G)$ is not empty by condition 4. We denote by $\psi_{\mathfrak{p}j}\colon \mathfrak{G}_\mathfrak{p} \to G_j$ the composite of $\psi_\mathfrak{p}$ with $G \to G_j$. Consider the diagrams

$$
\begin{tikzcd}
 & \mathfrak{G} \arrow[d, "\psi_{i-1}"] \\
G_i \arrow[r, "\pi_i"] & G_{i-1},
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \mathfrak{G}_\mathfrak{p} \arrow[dl, "\psi_{\mathfrak{p}i}"'] \arrow[d, "\psi_{i-1}|\mathfrak{G}_\mathfrak{p}"] \\
G_i \arrow[r] & G_{i-1},
\end{tikzcd}
\qquad \mathfrak{p} \in S_{i-1}.
$$

Since $[\psi_{\mathfrak{p}j}] \in \mathcal{H}om_{G_{i-1}}(\mathfrak{G}_\mathfrak{p}, G_i)$ the right diagrams are commutative. The kernel of π_i is the simple G_{i-1}-module A_i. Its exponent p_i divides n and so does the exponent n_i of the kernel H_i of $G \to G_i$, so that np_i is a multiple of $n_i p_i$ and $(np_i, m(K_{i-1})) = (n, m(K_{i-1})) = 1$ by condition 3. Since by condition 4

$$\prod_{\mathfrak{p}} \mathcal{H}om_{G_{i-1}}(\mathfrak{G}_{\mathfrak{p}}, G_i) \neq \emptyset$$

we can apply Lemma 6 to this situation replacing S by S_{i-1} and $E \to G$ by $G \to G_i$ and φ by ψ_{i-1}. We find an element $[\psi_i] \in \mathcal{H}om_{G_{i-1}}(\mathfrak{G}, G_i)_{\text{sur}}$ with the properties

(i) $[\psi_i | \mathfrak{G}_{\mathfrak{p}}] = [\psi_{\mathfrak{p}i}]$ in $\mathcal{H}om_{G_{i-1}}(\mathfrak{G}_{\mathfrak{p}}, G_i)$ for $\mathfrak{p} \in S_{i-1}$.
(ii) For $\mathfrak{p} \notin S_{i-1}$ we have $\mathcal{H}om_{G_i}(\mathfrak{G}_{\mathfrak{p}}, G) \neq \emptyset$.
(iii) For the field $K_i | k$ defined by ψ_i we have $(n, m(K_i)) = 1$.

Let ψ_i be a representative of $[\psi_i]$. Then clearly ψ_i is a surjective Γ-homomorphism $\mathfrak{G} \to G_i$, and we show that it satisfies the condition 1–4 of the claim.

Condition 1 is satisfied because of $[\psi_i] \in \mathcal{H}om_{G_{i-1}}(\mathfrak{G}, G_i)$. By (i) $\psi_i | \mathfrak{G}_{\mathfrak{p}}$ and $\psi_{\mathfrak{p}i}$ are Γ-homomorphisms, which are conjugate under an element of $A_i = H_i/H_{i-1} \subseteq H/H_{i-1} = \ker(G_i \to \Gamma)$, $\mathfrak{p} \in S$. Therefore $[\psi_i | \mathfrak{G}_{\mathfrak{p}}] = [\psi_{\mathfrak{p}i}]$ in $\mathcal{H}om_{\Gamma}(\mathfrak{G}_{\mathfrak{p}}, G_i)$, showing condition 2. Condition 3 is satisfied by (iii). Condition 4 finally follows from (ii) and from $[\psi_{\mathfrak{p}}] \in \mathcal{H}om_{G_i}(\mathfrak{G}_{\mathfrak{p}}, G)$ for $\mathfrak{p} \in S_{i-1}$.

The kernel of $G \to G_i$ is the group H_i. Since $\bigcap_i H_i = \{1\}$ we have $G = \varprojlim_i G_i$. Therefore the surjective Γ-homomorphisms $\psi_i \colon \mathfrak{G} \to G_i$ define a surjective Γ-homomorphism $\psi \colon \mathfrak{G} \to G$, i.e. an element $[\psi] \in \mathcal{H}om_{\Gamma}(\mathfrak{G}, G)_{\text{sur}}$. For each $\mathfrak{p} \in S$ we have $[\psi_i | \mathfrak{G}_{\mathfrak{p}}] = [\psi_{\mathfrak{p}i}]$ in $\mathcal{H}om_{\Gamma}(\mathfrak{G}_{\mathfrak{p}}, G_i)$; i.e.

$$\psi_i | \mathfrak{G}_{\mathfrak{p}} = a_{\mathfrak{p}i} \psi_{\mathfrak{p}i} a_{\mathfrak{p}i}^{-1}, \quad i = 1, 2, \ldots$$

with $a_{\mathfrak{p}i} \in \operatorname{Ker}(G_i \to \Gamma) = H/H_i$. A simple compactness argument now shows that we can find an element $\bar{a}_{\mathfrak{p}} \in H$ with

$$\psi_i | \mathfrak{G}_{\mathfrak{p}} = \bar{a}_{\mathfrak{p}i} \psi_{\mathfrak{p}i} \bar{a}_{\mathfrak{p}i}^{-1}, \quad i = 1, 2, \ldots$$

where $\bar{a}_{\mathfrak{p}i}$ is the image of $\bar{a}_{\mathfrak{p}}$ in H/H_i, and this means that $\psi | \mathfrak{G}_{\mathfrak{p}} = \bar{a}_{\mathfrak{p}} \psi_{\mathfrak{p}} \bar{a}_{\mathfrak{p}}^{-1}$, i.e. $[\psi | \mathfrak{G}_{\mathfrak{p}}] = [\psi_{\mathfrak{p}}]$ in $\mathcal{H}om_{\Gamma}(\mathfrak{G}, G)$.

This proves the Main Theorem.

§3. Applications of the Main Theorem

We now want to explain the meaning of the main theorem by giving some consequences and special cases. As before we shall assume that all occuring profinite groups are separable, i.e. have a countable basis of neighbourhoods.

Corollary 1. *Let k be a finite algebraic number field and G a pro-solvable group of finite odd exponent.*

Then there exists a Galois extension $K|k$ with Galois group isomorphic to G:

$$\operatorname{Gal}(K|k) \cong G.$$

Proof. We apply the main theorem to the field $\mathbb{Q}$ of rational numbers, i.e. we denote by $\mathfrak{G}$ resp. $\mathfrak{G}_p$ the absolute Galois group of $\mathbb{Q}$ resp. $\mathbb{Q}_p$. For S we choose the set of all primes of $\mathbb{Q}$ which ramify in k, and we set $\Gamma=\{1\}$. Then trivially $\prod\limits_p \mathcal{H}om_\Gamma(\mathfrak{G}_p, G)\neq\emptyset$ and the exponent of G is prime to $m(\mathbb{Q})=2$. Therefore by the main theorem there exists a surjective homomorphism $\psi\colon \mathfrak{G}\to G$, the restriction ψ_p of which to $\mathfrak{G}_p$ is the trivial homomorphism for $p\in S$. If K' is the fixed field of $\ker(\psi)$, then $\mathrm{Gal}(K'|\mathbb{Q})\cong\mathfrak{G}/\ker(\psi)\cong G$. Since $\mathfrak{G}_p$ is a decomposition group of the algebraic closure $\bar{\mathbb{Q}}$ of $\mathbb{Q}$ over p, the image of $\mathfrak{G}_p$ under $\mathfrak{G}\to G(K'|\mathbb{Q})$ is a decomposition group of $K'|\mathbb{Q}$ over p. Since $\psi(\mathfrak{G}_p)=\{1\}$ for $p\in S$, the primes in S split completely in K'. Therefore the field $K'\cap k$ is unramified over $\mathbb{Q}$, i.e. $K'\cap k=\mathbb{Q}$. If we set $K=K'\cdot k$ then $\mathrm{Gal}(K|k)\cong\mathrm{Gal}(K'|k)\cong G$.

We remark, that it is necessary to require the finiteness of the exponent of G. The group $\mathbb{Z}_p\times\mathbb{Z}_p$ for example cannot occur as a Galois group over $\mathbb{Q}$.

As a special case of the main theorem one obtains also the following existence theorem of Grunwald type.

Corollary 2. *Let k be a finite algebraic number field, $m(k)$ the number of its roots of unity and S a finite set of primes of k.*

Let G be a pro-solvable group of finite exponent prime to $m(k)$ and $K_\mathfrak{p}|k_\mathfrak{p}$, $\mathfrak{p}\in S$, Galois extensions whose Galois groups $G_\mathfrak{p}$ are imbeddable into G.

Then there exists a Galois extension $K|k$ with Galois group isomorphic to G, which for the primes $\mathfrak{p}\in S$ has the given extensions $K_\mathfrak{p}|k_\mathfrak{p}$ as completions.

Proof. We apply the main theorem to the case $\Gamma=1$. Then trivially

$$\prod_\mathfrak{p}\mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)\neq\emptyset,$$

and the field which in the theorem is denoted by K is identical with k. For every $\mathfrak{p}\in S$ let $\mathfrak{G}_\mathfrak{p}\to G_\mathfrak{p}$ be the canonical homomorphism onto the Galois group of $K_\mathfrak{p}|k_\mathfrak{p}$ and $G_\mathfrak{p}\to G$ an imbedding. Then we obtain homomorphisms $\psi_\mathfrak{p}$: $\mathfrak{G}_\mathfrak{p}\to G$ and by the main theorem there exists a surjective homomorphism ψ: $\mathfrak{G}\to G$ such that

$$\psi|\mathfrak{G}_\mathfrak{p}=a_\mathfrak{p}\,\psi_\mathfrak{p}\,a_\mathfrak{p}^{-1}\qquad\text{for }\ \mathfrak{p}\in S,$$

where $a_\mathfrak{p}\in G$. If $K|k$ is the fixed field of $\ker(\psi)$ then on the one hand $\mathrm{Gal}(K|k)\cong\mathfrak{G}/\ker(\psi)\cong G$. On the other hand the completion of $K|k$ at the prime $\mathfrak{p}\in S$ is the fixed field of the kernel of the restriction $\psi|\mathfrak{G}_\mathfrak{p}\colon\mathfrak{G}_\mathfrak{p}\to G$ and this kernel is the same as the kernel of $\psi_\mathfrak{p}$ which defines the given extension $K_\mathfrak{p}|k_\mathfrak{p}$.

The field theoretical content of the main theorem is a general assertion about the solvability of the so called *imbedding problem.* In order to explain this we consider the diagram

$$\begin{tikzcd}
 & \mathfrak{G} \arrow[d, "\varphi"] \\
G \arrow[r, "f"] & \Gamma
\end{tikzcd}$$

with a finite group Γ and surjective homomorphisms φ and f. We denote this diagram by $\mathscr{E}_\Gamma(\mathfrak{G}, G)$ and call it an imbedding problem. The homomorphism φ

defines a Galois extension $K|k$ with Galois group $\mathrm{Gal}(K|k) \cong \mathfrak{G}/\ker(\varphi) \cong \Gamma$. The existence of a surjective homomorphism $\psi: \mathfrak{G} \to G$ with $f \circ \psi = \varphi$ means accordingly the existence of a Galois extension $N \supseteq K \supseteq k$ with Galois group $\mathrm{Gal}(N|k)$ isomorphic to G such that the canonical projection

$$\mathrm{Gal}(N|k) \to \mathrm{Gal}(K|k)$$

coincides with the given homomorphism

$$G \xrightarrow{f} \Gamma$$

after the groups Γ and $\mathrm{Gal}(K|k)$ resp. G and $\mathrm{Gal}(N|k)$ have been identified. If ψ is not surjective then one obtains instead of the field N only a Galois algebra with Galois group G (or a field $N|k$ with Galois group $\psi(\mathfrak{G}) = \bar{G} \subseteq G$).

Thus the imbedding problem poses the task to imbed a given extension by a Γ-homomorphism $\psi: \mathfrak{G} \to G$ into a larger extension $N|k$ with a given Galois group. The equivalence classes of such homomorphisms ψ, i.e. the elements $[\psi]$ of $\mathcal{H}om_\Gamma(\mathfrak{G}, G)$ are called the *solutions* of the imbedding problem $\mathcal{E}_\Gamma(\mathfrak{G}, G)$. Note that two equivalent Γ-homomorphisms ψ define the same field N. The surjective elements $[\psi] \in \mathcal{H}om_\Gamma(\mathfrak{G}, G)_{\mathrm{sur}}$ are called *proper solutions*. They yield fields $N|k$ with Galois group G.

To every imbedding problem $\mathcal{E}_\Gamma(\mathfrak{G}, G)$ there are associated local imbedding problems over the completions $k_\mathfrak{p}$:

$$\begin{array}{ccccccc} & & \mathfrak{G} & & & & \mathfrak{G}_\mathfrak{p} \\ & & \Big\downarrow{\scriptstyle\varphi} & \mathcal{E}_\Gamma(\mathfrak{G}, G) & & & \Big\downarrow{\scriptstyle\varphi_\mathfrak{p}} & \mathcal{E}_{\Gamma_\mathfrak{p}}(\mathfrak{G}_\mathfrak{p}, G_\mathfrak{p}). \\ G & \xrightarrow{f} & \Gamma & & G_\mathfrak{p} & \xrightarrow{f_\mathfrak{p}} & \Gamma_\mathfrak{p} \end{array}$$

Here $\Gamma_\mathfrak{p} = \varphi(\mathfrak{G}_\mathfrak{p}) \subseteq \Gamma$, $G_\mathfrak{p} = f^{-1}(\Gamma_\mathfrak{p}) \subseteq G$ and $\varphi = \varphi|_{\mathfrak{G}_\mathfrak{p}}$, $f_\mathfrak{p} = f|_{\mathfrak{G}_\mathfrak{p}}$. The extension $K_\mathfrak{p}|k_\mathfrak{p}$ with Galois group $\Gamma_\mathfrak{p}$ defined by $\varphi_\mathfrak{p}: \mathfrak{G}_\mathfrak{p} \to \Gamma_\mathfrak{p}$ is just the completion at $\mathfrak{p}$ of the extension $K|k$ with Galois group Γ defined by $\varphi: \mathfrak{G} \to \Gamma$.

A solution $[\psi] \in \mathcal{H}om_\Gamma(\mathfrak{G}, G)$ of $\mathcal{E}_\Gamma(\mathfrak{G}, G)$ induces for every $\mathfrak{p}$ a solution $[\psi_\mathfrak{p}] \in \mathcal{H}om_{\Gamma_\mathfrak{p}}(\mathfrak{G}_\mathfrak{p}, G_\mathfrak{p}) = \mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)$ of $\mathcal{E}_{\Gamma_\mathfrak{p}}(\mathfrak{G}_\mathfrak{p}, G_\mathfrak{p})$, and the field $N_\mathfrak{p}(\supseteq K_\mathfrak{p} \supseteq k_\mathfrak{p})$ defined by $\psi_\mathfrak{p}$ is again the completion of the field $N(\supseteq K \supseteq k)$ defined by ψ. Thus the canonical map

$$\mathcal{H}om_\Gamma(\mathfrak{G}, G) \to \prod_\mathfrak{p} \mathcal{H}om_\Gamma(\mathfrak{G}_\mathfrak{p}, G)$$

has to be understood as a map between the sets of solutions of the imbedding problems $\mathcal{E}_\Gamma(\mathfrak{G}, G)$ and $\mathcal{E}_{\Gamma_\mathfrak{p}}(\mathfrak{G}_\mathfrak{p}, G_\mathfrak{p})$.

In the sequel we consider the imbedding problem under the assumption that the kernel of

$$G \xrightarrow{f} \Gamma$$

is a pro-solvable group of finite exponent prime to $m(K)$, where $m(K)$ is the number of roots of unity in the field K defined by φ. If for example K is a real number field we consider imbedding problems where the kernel of $G \xrightarrow{f} \Gamma$ is a prosolvable group of finite odd exponent.

We have now the following local-global-principle.

Corollary 3. *The imbedding problem $\mathscr{E}_\Gamma(\mathfrak{G}, G)$ has a proper solution if and only if all the local imbedding problems $\mathscr{E}_{\Gamma_\mathfrak{p}}(\mathfrak{G}_\mathfrak{p}, G_\mathfrak{p})$ are solvable.*

Proof. The solvability of all local imbedding problems $\mathscr{E}_{\Gamma_\mathfrak{p}}(\mathfrak{G}_\mathfrak{p}, G_\mathfrak{p})$ is equivalent with $\prod_\mathfrak{p} \mathscr{H}\!om_\Gamma(\mathfrak{G}_\mathfrak{p}, G) \neq \emptyset$. Therefore by the main theorem the set $\mathscr{H}\!om_\Gamma(\mathfrak{G}, G)_{\mathrm{sur}}$ is not empty.

This corollary can be understood as a complete answer to the imbedding problem in so far, as it reduces the solvability question to a purely group theoretical problem. Namely for the primes $\mathfrak{p}$ which are unramified in $K|k$ the imbedding problem $\mathscr{E}_{\Gamma_\mathfrak{p}}(\mathfrak{G}_\mathfrak{p}, G_\mathfrak{p})$ is automatically solvable, so that it only remains to check the solvability at the finitely many ramified primes $\mathfrak{p}$. Now by the work of Jakovlev [6] the structure of the group $\mathfrak{G}_\mathfrak{p}$ is explicitly known. It is given in terms of generators and relations. Therefore the problem of finding a $\Gamma_\mathfrak{p}$-homomorphism $\mathfrak{G}_\mathfrak{p} \to G_\mathfrak{p}$ goes over into a purely group theoretical word problem. In [9], Korollar 9, 10, 11 we have given examples, which demonstrate, how this group theoretical question can be decided in concrete cases.

Since by Lemma 5 the local problem $\mathscr{E}_{\Gamma_\mathfrak{p}}(\mathfrak{G}_\mathfrak{p}, G_\mathfrak{p})$ is always solvable if the extension $K_\mathfrak{p}|k_\mathfrak{p}$ is unramified we obtain the result

Corollary 4. *The imbedding problem $\mathscr{E}_\Gamma(\mathfrak{G}, G)$ has always a proper solution if the extension $K|k$ is unramified.*

Directly from the main theorem follows furthermore

Corollary 5. *If the imbedding problem $\mathscr{E}_\Gamma(\mathfrak{G}, G)$ has a solution, then it has also a proper solution.*

Proof. $\mathscr{H}\!om_\Gamma(\mathfrak{G}, G) \neq \emptyset$ implies $\prod_\mathfrak{p} \mathscr{H}\!om_\Gamma(\mathfrak{G}_\mathfrak{p}, G) \neq \emptyset$ and hence $\mathscr{H}\!om_\Gamma(\mathfrak{G}, G)_{\mathrm{sur}} \neq \emptyset$ by the main theorem.

The assertion of this Corollary has been proven by Ikeda in the case that the kernel $G \xrightarrow{f} \Gamma$ is a finite abelian group.

Corollary 6. *The imbedding problem $\mathscr{E}_\Gamma(\mathfrak{G}, G)$ has always a proper solution if the group extension $G \xrightarrow{f} \Gamma$ splits.*

Proof. If $s: \Gamma \to G$ is a homomorphism with $f \circ s = \mathrm{id}_\Gamma$ then $\psi = s \circ \varphi: \mathfrak{G} \to G$ defines a solution of $\mathscr{E}_\Gamma(\mathfrak{G}, G)$. Hence by Corollary 5 there exists a proper solution.

This result has been proven by Scholz [16] in the case that the kernel of $G \to \Gamma$ is a finite abelian group and by Šafarevič [15] and Ishanov [5] in the case that the kernel is nilpotent.

We finally mention the following cohomological result.

Corollary 7. *Let A be a finite $\mathfrak{G}$-module such that $(\# A, m(k(A)))=1$. Then for every finite set S of primes of k the homomorphism*

$$H^1(\mathfrak{G}, A) \to \prod_{\mathfrak{p}\in S} H^1(\mathfrak{G}_{\mathfrak{p}}, A)$$

is surjective.

Proof. Let G be the semi-direct product of $\Gamma = \mathrm{Gal}(k(A)|k)$ with A.

By the proposition of §1 the pair

$$\mathscr{H}om_\Gamma(\mathfrak{G}, G) \to \prod_{\mathfrak{p}\in S} \mathscr{H}om_\Gamma(\mathfrak{G}_{\mathfrak{p}}, G), \qquad H^1(\mathfrak{G}, A) \to \prod_{\mathfrak{p}\in S} H^1(\mathfrak{G}_{\mathfrak{p}}, A)$$

is a morphism of principal homogeneous spaces. Since by the main theorem the left map is surjective, so is the right map.

The corollary applies especially to the case that A is of odd order and $k(A)$ is real. The homomorphism $H^1(\mathfrak{G}, A) \to \prod_{\mathfrak{p}\in S} H^1(\mathfrak{G}_{\mathfrak{p}}, A)$ however is in general not surjective as was shown in [9], Satz (6.5). This means at the same time that the main theorem becomes false if we drop the condition that the exponent of the kernel of $G \xrightarrow{f} \Gamma$ is prime to $m(K)$.

Appendix

Existence of Algebraic Numbers with Certain Local Properties

Our aim in this appendix is to establish the existence of certain numbers α in an algebraic number field, which we have used in the proof of the main lemma of §1. For this we shall make use of an idea of Šafarevič, who proved a similar existence theorem in [14].

Let $K|k$ be a Galois extension of finite algebraic number fields and p a fixed odd prime number. Without further mention we assume in the sequel that K contains the p-th roots of unity. If $\mathfrak{p}$ is a prime of K then $K_{\mathfrak{p}}$ denotes the completion of K at $\mathfrak{p}$ and $U_{\mathfrak{p}}$ the group of units in $K_{\mathfrak{p}}$. If α is an element of K^* or K^*/K^{*p} then we denote by $\alpha_{\mathfrak{p}}$ the image of α in $K_{\mathfrak{p}}^*/K_{\mathfrak{p}}^{*p}$. Sometimes however we write also simply α instead of $\alpha_{\mathfrak{p}}$, if there is no danger of confusion.

If $\Omega|K$ is an abelian extension of exponent p, then the map

$$(\ ,\Omega|K): \prod_{\mathfrak{p}} K_{\mathfrak{p}}^*/K_{\mathfrak{p}}^{*p} \to \mathrm{Gal}(\Omega|K)$$

is defined by the converging product

$$(a, \Omega|K) = \prod_{\mathfrak{p}} (a_{\mathfrak{p}}, \Omega_{\mathfrak{p}}|K_{\mathfrak{p}})$$

for $a = (a_{\mathfrak{p}}) \in \prod_{\mathfrak{p}} K_{\mathfrak{p}}^*/K_{\mathfrak{p}}^{*p}$, where $\Omega_{\mathfrak{p}}|K_{\mathfrak{p}}$ is the completion of $\Omega|K$ and

$$(\ ,\Omega_{\mathfrak{p}}\,|\,K_{\mathfrak{p}})\colon K_{\mathfrak{p}}^*/K_{\mathfrak{p}}^{*p}\to \mathrm{Gal}(\Omega_{\mathfrak{p}}\,|\,K_{\mathfrak{p}})\subseteq \mathrm{Gal}(\Omega\,|\,K)$$

is the local norm residue symbol. We shall make repeated use of the fact that $(a_{\mathfrak{p}},\Omega_{\mathfrak{p}}\,|\,K_{\mathfrak{p}})=1$ if $\Omega_{\mathfrak{p}}\,|\,K_{\mathfrak{p}}$ is unramified and $a_{\mathfrak{p}}\in U_{\mathfrak{p}}K_{\mathfrak{p}}^{*p}/K_{\mathfrak{p}}^{*p}$.

Theorem. *Let $K\,|\,k$ and $\Omega\,|\,K$ be two Galois extensions and let $\Omega_0\,|\,K$ be the maximal abelian subextension of exponent p of $\Omega\,|\,K$.*

Let S be a finite set of primes of K and $\beta_{\mathfrak{p}}\in K_{\mathfrak{p}}^{*p}$ for $\mathfrak{p}\in S$ such that

$$\prod_{\mathfrak{p}\in S}(\beta_{\mathfrak{p}},\Omega_{0\mathfrak{p}}\,|\,K_{\mathfrak{p}})=1.$$

Then there exists an element $\alpha\in K^*/K^{*p}$ with the properties

(i) $\alpha_{\mathfrak{p}}=\beta_{\mathfrak{p}}$ *for* $\mathfrak{p}\in S$.

(ii) *If $\alpha_{\mathfrak{p}}\notin U_{\mathfrak{p}}K_{\mathfrak{p}}^{*p}/K_{\mathfrak{p}}^{*p}$ for $\mathfrak{p}\notin S$, then the prime $\mathfrak{p}_0=\mathfrak{p}\cap k$ of k lying under $\mathfrak{p}$ splits completely in Ω and $\alpha_{\mathfrak{p}'}=1$ for the primes $\mathfrak{p}'\,|\,\mathfrak{p}_0$ of K different from $\mathfrak{p}$.*

Proof. We may assume that the set S contains all primes above p, all infinite primes and all primes of K which ramify in Ω, by completing the set of $\beta_{\mathfrak{p}}$, $\mathfrak{p}\in S$, eventually by $\beta_{\mathfrak{p}}=1$. We prove the theorem first in the case that

$$\beta_{\mathfrak{p}}\in U_{\mathfrak{p}}K_{\mathfrak{p}}^{*p}/K_{\mathfrak{p}}^{*p}\quad\text{for all }\ \mathfrak{p}\in S\quad\text{and}\quad \beta_{\mathfrak{p}}=1\quad\text{for }\mathfrak{p}\,|\,p.$$

The number α of K which we are looking for will be obtained as a product of two members of a sequence

$$\pi_1,\pi_2,\pi_3,\ldots$$

of elements of K, such that for each i we have

$$(\pi_i)=\mathfrak{p}_i\,\mathfrak{a}_i^p$$

with a prime ideal $\mathfrak{p}_i$ which is not conjugate over k to any of the prime ideals $\mathfrak{p}_j$, $j\neq i$, and $\mathfrak{p}\in S$. If such a sequence is given we set

$$S_n=S\cup\{\sigma\mathfrak{p}_i\,|\,i=1,\ldots,n;\sigma\in G-\{1\}\}$$

$K_n=$ the maximal abelian extension of exponent p of K, which is unramified outside S_n.

We shall construct the sequence $\pi_1,\pi_2,\pi_3,\ldots$ recursively in such a way that it satisfies the conditions

(1) $\mathfrak{p}_i\cap k$ splits completely in Ω.

(2) $\pi_{i\mathfrak{p}}=\beta_{\mathfrak{p}}^{1/2}$ for $\mathfrak{p}\in S$ and all i,

(3) $\left(\dfrac{\pi_{n+1}}{\sigma\mathfrak{p}_i}\right)=\left(\dfrac{\pi_i}{\sigma\mathfrak{p}_i}\right)^{-1}$ for $i=1,\ldots,n$, $\sigma\in G-\{1\}$,

(4) $\left(\dfrac{\pi_i}{\sigma\mathfrak{p}_{n+1}}\right)=\left(\dfrac{\pi_i}{\sigma\mathfrak{p}_i}\right)^{-1}$ for $i=1,\ldots,n$, $\sigma\in G-\{1\}$,

where $\left(\dfrac{a}{\mathfrak{a}}\right)$ denotes the p-th power residue symbol.

Let us assume that we have already constructed the elements $\pi_1, \ldots, \pi_n$. Let $u_\mathfrak{p} = \beta_\mathfrak{p}^{1/2} \in U_\mathfrak{p} K_\mathfrak{p}^{*p}/K_\mathfrak{p}^{*p}$ for $\mathfrak{p} \in S$ (notice that the exponent 1/2 has a meaning since $p \neq 2$) and let us consider the automorphism

$$\tau = \prod_{\mathfrak{p} \in S} (u_\mathfrak{p}, K_{n\mathfrak{p}} | K_\mathfrak{p}) \prod_{i=1}^{n} \prod_{\sigma \neq 1} (\pi_i, K_{n\sigma\mathfrak{p}_i} | K_{\sigma\mathfrak{p}_i})^{-1} \in \mathrm{Gal}(K_n | K). \tag{5}$$

Since S_n contains the primes which are ramified in Ω we have $\Omega_0 \subseteq K_n$, i.e. $\Omega_0 = \Omega \cap K_n$. Since $\prod_{\mathfrak{p} \in S} (u_\mathfrak{p}, \Omega_{0\mathfrak{p}} | K_\mathfrak{p}) = 1$ by assumption, since moreover $\Omega_{0\sigma\mathfrak{p}_i} | K_{\sigma\mathfrak{p}_i}$ is unramified and π_i is a unit in $K_{\sigma\mathfrak{p}_i}$ we have $\tau|_{\Omega_0} = 1$. Therefore τ can be lifted to an automorphism $\tilde{\tau}$ of $\Omega_n = \Omega K_n$ which is the identity on Ω, i.e. $\tilde{\tau} \in \mathrm{Gal}(\Omega_n | \Omega)$ and $\tilde{\tau}|_{K_n} = \tau$. Now by Čebotarev's density theorem there exists a prime $\mathfrak{q}$ of K with the Artin symbol

$$\left[\frac{\Omega_n | K}{\mathfrak{q}}\right] = \langle \tilde{\tau}^{-1} \rangle,$$

$\langle \tilde{\tau}^{-1} \rangle$ being the conjugacy class of $\tilde{\tau}^{-1}$ in $\mathrm{Gal}(\Omega_n | K)$, and we can choose $\mathfrak{q}$ in such a way that $\mathfrak{q} \cap k$ splits completely in Ω and is not conjugate to any of the primes $\mathfrak{p} \in S$ and $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$. If $\pi_\mathfrak{q}$ is a prime element of $K_\mathfrak{q}$, we have

$$\tau = \tilde{\tau}|_{K_n} = \left(\frac{K_n | K}{\mathfrak{q}}\right)^{-1} = (\pi_\mathfrak{q}, K_{n\mathfrak{q}} | K_\mathfrak{q})^{-1}$$

and together with (5) we obtain

$$\prod_{\mathfrak{p} \in S} (u_\mathfrak{p}, K_{n\mathfrak{p}} | K_\mathfrak{p}) \prod_{i=1}^{n} \prod_{\sigma \neq 1} (\pi_i, K_{n\sigma\mathfrak{p}_i} | K_{\sigma\mathfrak{p}_i})^{-1} (\pi_\mathfrak{q}, K_{n\mathfrak{q}} | K_\mathfrak{q}) = 1. \tag{6}$$

Now by class field theory we have an exact sequence

$$K^*/K^{*p} \to \prod_{\mathfrak{p} \in S_n} K_\mathfrak{p}^*/K_\mathfrak{p}^{*p} \times \prod_{\mathfrak{p} \notin S_n} K_\mathfrak{p}^*/U_\mathfrak{p} K_\mathfrak{p}^{*p} \xrightarrow{(\ ,\, K_n | K)} \mathrm{Gal}(K_n | K) \to 1.$$

Because of (6) there exists an element $\pi_{n+1} \in K^*$ such that

(7) $\pi_{n+1,\mathfrak{p}} = u_\mathfrak{p}$ for $\mathfrak{p} \in S$,

(8) $\pi_{n+1,\sigma\mathfrak{p}_i} = \pi_{i\sigma\mathfrak{p}_i}^{-1}$ for $i = 1, \ldots, n; \sigma \in G - \{1\}$,

(9) $\pi_{n+1} \equiv \pi_\mathfrak{q} \bmod U_\mathfrak{q} K_\mathfrak{q}^{*p}$,

(10) $\pi_{n+1} \in U_\mathfrak{p} K_\mathfrak{p}^{*p}$ for $\mathfrak{p} \notin S_n \cup \{\mathfrak{q}\}$.

We now show that the element π_{n+1} satisfies the conditions (1)–(4). Since π_{n+1} is a unit times a p-th power in $K_\mathfrak{p}$ for all $\mathfrak{p} \neq \mathfrak{q}$ we have

$$(\pi_{n+1}) = \mathfrak{p}_{n+1} \mathfrak{a}_{n+1}^p \quad \text{with} \quad \mathfrak{p}_{n+1} = \mathfrak{q}.$$

Since $\mathfrak{q} \cap k$ splits completely in Ω condition (1) is satisfied. The conditions (2) and (3) are satisfied because of (7) and (8). It remains to prove condition (4). Let $i \in \{1, \ldots, n\}$ and $\sigma \in G - \{1\}$. The extension $K(\sqrt[p]{\sigma\pi_i}) | K$ lies in K_n since it is ramified only at $\sigma\mathfrak{p}_i$ and eventually at the primes $\mathfrak{p}|p$. Therefore we have

$$\tau|_{K(\sqrt[p]{\sigma\pi_i})}=\left(\frac{K(\sqrt[p]{\sigma\pi_i})|K}{\mathfrak{p}_{n+1}}\right)^{-1}$$

and on the other hand

$$\tau|_{K(\sqrt[p]{\sigma\pi_i})}=\prod_{\mathfrak{p}\in S}(u_\mathfrak{p}, K_\mathfrak{p}(\sqrt[p]{\sigma\pi_i})|K_\mathfrak{p})\cdot\prod_{j=1}^{n}\prod_{\rho\neq 1}(\pi_j, K_{\rho\mathfrak{p}_j}(\sqrt[p]{\sigma\pi_i})|K_{\rho\mathfrak{p}_j})^{-1}.$$

The extension $K_\mathfrak{p}(\sqrt[p]{\sigma\pi_i})|K_\mathfrak{p}$ is ramified at most at the primes $\mathfrak{p}=\sigma\mathfrak{p}_i$ and $\mathfrak{p}|p$. Since $u_\mathfrak{p}=1$ for $\mathfrak{p}|p$ and since the $u_\mathfrak{p}$, π_j are local units we have

$$\tau|_{K(\sqrt[p]{\sigma\pi_i})}=(\pi_i, K_{\sigma\mathfrak{p}_i}(\sqrt[p]{\sigma\pi_i})|K_{\sigma\mathfrak{p}_i})^{-1}.$$

Using the product formula for the local norm residue symbols we get

$$\tau|_{K(\sqrt[p]{\sigma\pi_i})}=\prod_{\mathfrak{p}\neq\sigma\mathfrak{p}_i}(\pi_i, K_\mathfrak{p}(\sqrt[p]{\sigma\pi_i})|K_\mathfrak{p})=(\pi_i, K_{\mathfrak{p}_i}(\sqrt[p]{\sigma\pi_i})|K_{\mathfrak{p}_i})=\left(\frac{K(\sqrt[p]{\sigma\pi_i})|K}{\mathfrak{p}_i}\right)$$

and hence

$$\left(\frac{K(\sqrt[p]{\sigma\pi_i})|K}{\mathfrak{p}_{n+1}}\right)=\left(\frac{K(\sqrt[p]{\sigma\pi_i})|K}{\mathfrak{p}_i}\right)^{-1}.$$

Applying this equality to $\sqrt[p]{\sigma\pi_i}$ we obtain $\left(\frac{\sigma\pi_i}{\mathfrak{p}_{n+1}}\right)=\left(\frac{\sigma\pi_i}{\mathfrak{p}_i}\right)^{-1}$ and therefore $\left(\frac{\pi_i}{\sigma^{-1}\mathfrak{p}_{n+1}}\right)=\left(\frac{\pi_i}{\sigma^{-1}\mathfrak{p}_i}\right)^{-1}$. This proves condition (4).

Let $G-\{1\}=\{\sigma_1,\dots,\sigma_r\}$. Then for any i the r-tuple of power residue symbols $\left(\left(\frac{\pi_i}{\sigma_1\mathfrak{p}_i}\right),\dots,\left(\frac{\pi_i}{\sigma_r\mathfrak{p}_i}\right)\right)$ is an element of the finite set μ_p^r. Therefore by the shoe box principle we can find two numbers $s<t$ such that

$$\left(\frac{\pi_s}{\sigma\mathfrak{p}_s}\right)=\left(\frac{\pi_t}{\sigma\mathfrak{p}_t}\right)\quad\text{for all } \sigma\in G-\{1\}.$$

We now prove that the element $\alpha=\pi_s\cdot\pi_t$ satisfies the conditions of the theorem.

The condition (1) is clear since $\alpha_\mathfrak{p}=\pi_{s\mathfrak{p}}\pi_{t\mathfrak{p}}=\beta_\mathfrak{p}^{1/2}\beta_\mathfrak{p}^{1/2}=\beta_\mathfrak{p}$ for $\mathfrak{p}\in S$. Assume that $\mathfrak{p}\notin S$ is a prime such that $\alpha_\mathfrak{p}\notin U_\mathfrak{p}K_\mathfrak{p}^{*p}/K_\mathfrak{p}^{*p}$. Then either $\mathfrak{p}=\mathfrak{p}_s$ or $\mathfrak{p}=\mathfrak{p}_t$. In both cases $\mathfrak{p}\cap k$ splits completely in Ω by (1). In the case $\mathfrak{p}=\mathfrak{p}_s$ we have by (3)

$$\left(\frac{\alpha}{\sigma\mathfrak{p}_s}\right)=\left(\frac{\pi_s}{\sigma\mathfrak{p}_s}\right)\left(\frac{\pi_t}{\sigma\mathfrak{p}_s}\right)=\left(\frac{\pi_s}{\sigma\mathfrak{p}_s}\right)\left(\frac{\pi_s}{\sigma\mathfrak{p}_s}\right)^{-1}=1\quad\text{for } \sigma\in G-\{1\}$$

and hence $\alpha\in K_{\sigma\mathfrak{p}_s}^{*p}$, i.e. $\alpha_{\sigma\mathfrak{p}}=1$. In the case $\mathfrak{p}=\mathfrak{p}_t$ we have

$$\left(\frac{\alpha}{\sigma\mathfrak{p}_t}\right)=\left(\frac{\pi_s}{\sigma\mathfrak{p}_t}\right)\left(\frac{\pi_t}{\sigma\mathfrak{p}_t}\right)=\left(\frac{\pi_s}{\sigma\mathfrak{p}_s}\right)^{-1}\left(\frac{\pi_s}{\sigma\mathfrak{p}_s}\right)=1\quad\text{for } \sigma\in G-\{1\}$$

and hence also $\alpha_{\sigma\mathfrak{p}}=1$. This proves the theorem under the assumption $\beta_{\mathfrak{p}}\in U_{\mathfrak{p}}K_{\mathfrak{p}}^{*p}/K_{\mathfrak{p}}^{*p}$ for $\mathfrak{p}\in S$ and $\beta_{\mathfrak{p}}=1$ for $\mathfrak{p}\,|\,p$.

We now prove the theorem in the general case. Let $K^S\,|\,K$ be the maximal abelian extension of exponent p which is unramified outside S. Since $\Omega\,|\,K$ is also unramified outside S we have $\Omega_0=\Omega\cap K^S$. Since the element $\tau=\prod_{\mathfrak{p}\in S}(\beta_{\mathfrak{p}}, K_{\mathfrak{p}}^S\,|\,K_{\mathfrak{p}})$ is the identity on Ω_0, it can be lifted to an automorphism $\tilde{\tau}$ of ΩK^S, which is the identity on Ω. By Čebotarev's density theorem we can choose a prime ideal $\mathfrak{q}\notin S$ of degree one over k, which splits completely in Ω and has Artin symbol

$$\left[\frac{\Omega K^S\,|\,K}{\mathfrak{q}}\right]=\langle\tilde{\tau}^{-1}\rangle.$$

Let $\pi_{\mathfrak{q}}$ be a prime element in $K_{\mathfrak{q}}$. By class field theory we have an exact sequence

$$K^*/K^{*p}\to I_S\xrightarrow{(\ ,\,K^S|K)}\mathrm{Gal}(K^S\,|\,K)\to 1,$$

where $I_S=\prod_{\mathfrak{p}\in S}K_{\mathfrak{p}}^*/K_{\mathfrak{p}}^{*p}\times\prod_{\mathfrak{p}\notin S}K_{\mathfrak{p}}^*/U_{\mathfrak{p}}K_{\mathfrak{p}}^{*p}$. If we define the element $x=(x_{\mathfrak{p}})\in I_S$ by

$$x_{\mathfrak{p}}=\beta_{\mathfrak{p}}\quad\text{for }\mathfrak{p}\in S,$$

$$x_{\mathfrak{q}}=\pi_{\mathfrak{q}}U_{\mathfrak{q}}K_{\mathfrak{q}}^{*p},$$

$$x_{\mathfrak{p}}=1\quad\text{for }\mathfrak{p}\notin S\cup\{\mathfrak{q}\},$$

then

$$(x, K^S\,|\,K)=\prod_{\mathfrak{p}}(x_{\mathfrak{p}}, K_{\mathfrak{p}}^S\,|\,K_{\mathfrak{p}})=\prod_{\mathfrak{p}\in S}(\beta_{\mathfrak{p}}, K_{\mathfrak{p}}^S\,|\,K_{\mathfrak{p}})\cdot(\pi_{\mathfrak{q}}, K_{\mathfrak{q}}^S\,|\,K_{\mathfrak{q}})=\tau\cdot\tau^{-1}=1.$$

Therefore there exists an element $\delta\in K^*/K^{*p}$ such that

$$\delta_{\mathfrak{p}}=\beta_{\mathfrak{p}}\ \text{ for }\mathfrak{p}\in S,\qquad \delta_{\mathfrak{p}}\in U_{\mathfrak{p}}K_{\mathfrak{p}}^{*p}/K_{\mathfrak{p}}^{*p}\ \text{ for }\mathfrak{p}\notin S\cup\{\mathfrak{q}\}.$$

Let $\Omega'=\Omega(\sqrt[p]{\delta})$ and $S'=S\cup\{\mathfrak{p}\,|\,\mathfrak{q}_0\}$, where $\mathfrak{q}_0=\mathfrak{q}\cap k$. The primes of K which are ramified in Ω' lie in the set $S\cup\{\mathfrak{q}\}$. We define the elements $u_{\mathfrak{p}}\in U_{\mathfrak{p}}K_{\mathfrak{p}}^{*p}/K_{\mathfrak{p}}^{*p}$ for $\mathfrak{p}\in S'$ by $u_{\mathfrak{p}}=1$ for $\mathfrak{p}\in S\cup\{\mathfrak{q}\}$, $u_{\mathfrak{p}}=\delta_{\mathfrak{p}}^{-1}$ for $\mathfrak{p}\,|\,\mathfrak{q}_0$, $\mathfrak{p}\neq\mathfrak{q}$. It is then clear that

$$\prod_{\mathfrak{p}\in S'}(u_{\mathfrak{p}}, \Omega'_{0\mathfrak{p}}\,|\,K_{\mathfrak{p}})=1$$

since the maximal subextension $\Omega'_0\,|\,K$ of exponent p of $\Omega'\,|\,K$ is unramified for the primes $\mathfrak{p}\,|\,\mathfrak{q}_0$, $\mathfrak{p}\neq\mathfrak{q}$.

Therefore we can apply what we have proved before. We find an element $\gamma\in K^*/K^{*p}$ with the properties

(11) $\gamma_{\mathfrak{p}}=u_{\mathfrak{p}}$ for $\mathfrak{p}\in S'$.

(12) If $\gamma_{\mathfrak{p}}\notin U_{\mathfrak{p}}K_{\mathfrak{p}}^{*p}$ then $\mathfrak{p}_0=\mathfrak{p}\cap k$ splits completely in Ω' and $\gamma_{\mathfrak{p}'}=1$ for all $\mathfrak{p}'\,|\,\mathfrak{p}_0$, $\mathfrak{p}'\neq\mathfrak{p}$.

We now show that the element $\alpha = \gamma \cdot \delta$ satisfies the conditions of the theorem. For $\mathfrak{p} \in S$ we have

$$\alpha_{\mathfrak{p}} = \gamma_{\mathfrak{p}} \cdot \delta_{\mathfrak{p}} = u_{\mathfrak{p}} \cdot \beta_{\mathfrak{p}} = \beta_{\mathfrak{p}}.$$

Let $\mathfrak{p} \notin S$ and assume that $\alpha_{\mathfrak{p}} = \gamma_{\mathfrak{p}} \, \delta_{\mathfrak{p}} \notin U_{\mathfrak{p}} \, K_{\mathfrak{p}}^{*p}/K_{\mathfrak{p}}^{*p}$. If $\mathfrak{p} \in S' - S$ then necessarily $\mathfrak{p} = \mathfrak{q}$. Since $\mathfrak{q}$ is of degree one over k and splits completely in Ω, the prime $\mathfrak{q}_0 = \mathfrak{q} \cap k$ of k splits also completely in Ω. If $\mathfrak{q}' | \mathfrak{q}_0$ and $\mathfrak{q}' \neq \mathfrak{q}$ then

$$\alpha_{\mathfrak{q}'} = \gamma_{\mathfrak{q}'} \, \delta_{\mathfrak{q}'} = u_{\mathfrak{q}'} \, u_{\mathfrak{q}'}^{-1} = 1.$$

If $\mathfrak{p} \notin S'$ then $\delta_{\mathfrak{p}} \in U_{\mathfrak{p}} \, K_{\mathfrak{p}}^{*p}/K_{\mathfrak{p}}^{*p}$ and therefore $\gamma_{\mathfrak{p}} \notin U_{\mathfrak{p}} \, K_{\mathfrak{p}}^{*p}/K_{\mathfrak{p}}^{*p}$. By (11) this means that $\mathfrak{p}_0 = \mathfrak{p} \cap k$ splits completely in Ω', and $\gamma_{\mathfrak{p}'} = 1$ for all $\mathfrak{p}' | \mathfrak{p}_0$, $\mathfrak{p}' \neq \mathfrak{p}$. On the other hand we have $\delta \in K_{\mathfrak{p}'}^{*p}$ since $\mathfrak{p}'$ splits completely in $K(\sqrt[p]{\delta}) \subseteq \Omega'$, so that $\delta_{\mathfrak{p}'} = 1$ and thus $\alpha_{\mathfrak{p}'} = \gamma_{\mathfrak{p}'} \, \delta_{\mathfrak{p}'} = 1$.

References

1. Bauer, M.: Zur Theorie der algebraischen Zahlkörper. Math. Ann. **77**, 353–356 (1916)
2. Hoechsmann, K.: Zum Einbettungsproblem. J. reine u. angew. Math. **229**, 81–106 (1968)
3. Ikeda, M.: Zur Existenz eigentlicher galoisscher Körper beim Einbettungsproblem. Hamb. Abh. **24**, 126–131 (1960)
4. Ishanov, V.V.: The imbedding problem with given localizations. Izv. Akad. Nauk. SSSR. Ser. Math. **39**, (1975) (russisch)
5. Ishanov, V.V.: On the semidirect imbedding problem with nilpotent kernel. Izv. Akad. Nauk. SSSR. **40**, (1976) (russisch)
6. Jakovlev, A.V.: The Galois group of the algebraic closure of a local field. Izv. Akad. Nauk. SSSR. **32**, (1968) (russisch)
7. Neukirch, J.: Einbettungsprobleme mit lokaler Vorgabe und freie Produkte lokaler Galoisgruppen. J. reine angew. Math. **259**, 1–47 (1973)
8. Neukirch, J.: Über das Einbettungsproblem der algebraischen Zahlentheorie. Inv. Math. **21**, 59–116 (1973)
9. Neukirch, J.: Zur Theorie der nilpotenten algebraischen Zahlkörper. Math. Ann. **219**, 65–83 (1976)
10. Neumann, O.: Einige Klassen von endlichen Galois-Moduln mit Hasse-Prinzip. Math. Nachr. **72**, 306–320 (1976)
11. Neumann, O.: Über das Einbettungsproblem für globale Körper bei begrenzter Verzweigung. Math. Nachr. **71**, 147–152 (1976)
12. Šafarevič, I.R.: On the construction of fields with a given galois group of order l^a. Izv. Akad. Nauk. SSSR. Ser. Math. **18**, 216–296 (1954), AMS Translation, Ser. 2, **4**, 107–142 (1956)
13. Šafarevič, I.R.: On the problem of imbedding fields. Jzv. Akad. Nauk. SSSR. Ser. Math. **18**, 389–418 (1954); AMS Translation, Ser. 2, **4**, 151–183 (1956)
14. Šafarevič, I.R.: Construction of fields of algebraic numbers with given solvable galois group. Izv. Akad. Nauk. SSSR. Ser. Math. **18**, 525–578 (1954); AMS Translation, Ser. 2, **4**, 185–237 (1956)
15. Šafarevič, I.R.: The imbedding problem for splitting extensions. Dokl. Akad. Nauk. SSSR. **120**, 1217–1219 (1958) (russisch)
16. Scholz, A.: Über die Bildung algebraischer Zahlkörper mit auflösbarer galoisscher Gruppe. Math. Z. **30**, 332–356 (1929)
17. Scholz, A.: Konstruktion algebraischer Zahlkörper mit beliebiger Gruppe von Primzahlpotenzordnung. Math. Z. **42**, 161–188 (1936)
18. Serre, J.P.: Cohomologie Galoisienne. Lecture Notes in Mathematics **5**, Berlin-Göttingen-Heidelberg: Springer 1964

Received January 20/April 9, 1979

Inventiones math. 53, 165–184 (1979)

Inventiones mathematicae

Representations of Coxeter Groups and Hecke Algebras

David Kazhdan[1] and George Lusztig[2]*

[1] Harvard University, Department of Mathematics, Cambridge, MA02138, USA
[2] MIT, Department of Mathematics, Cambridge, MA02139, USA

§1. Introduction

Let W be a Coxeter group and let S be the corresponding set of simple reflections. Following [2, Ch. IV, §2, Ex. 34], we define an algebra $\tilde{\mathscr{H}}$ over the polynomial ring $\mathbb{Z}[q]$ as follows. $\tilde{\mathscr{H}}$ has basis elements T_w, one for each $w \in W$. The multiplication is defined by the rules

$$T_w T_{w'} = T_{ww'}, \qquad \text{if } l(ww') = l(w) + l(w')$$
$$(T_s + 1)(T_s - q) = 0, \qquad \text{if } s \in S;$$

here $l(w)$ is the length of w.

In the case where W is a Weyl group and q is specialized to a fixed prime power, $\tilde{\mathscr{H}} \otimes_{\mathbb{Z}[q]} \mathbb{C}$ can be interpreted as the algebra of intertwining operators of the space of functions on the flag manifold of the corresponding finite Chevalley group $G(F_q)$ (see [loc. cit., Ex. 24]). Therefore, the problem of decomposing this space of functions into irreducible representations of $G(F_q)$ is equivalent to the problem of decomposing the regular representation of $\tilde{\mathscr{H}} \otimes_{\mathbb{Z}[q]} \mathbb{C}$. It is known that, in this case, $\tilde{\mathscr{H}} \otimes_{\mathbb{Z}[q]} \mathbb{C}$ is isomorphic to the group algebra of W; however, in general, this isomorphism cannot be defined without introducing a square root of q (see [1]).

It is therefore, natural to extend the ground ring of $\tilde{\mathscr{H}}$ as follows. For any Coxeter group (W, S) we define the Hecke algebra $\mathscr{H}$ to be $\tilde{\mathscr{H}} \otimes_{\mathbb{Z}[q]} A$, where A is the ring of Laurent polynomials with integral coefficients in the indeterminate $q^{1/2}$.

Our purpose is to construct representations of $\mathscr{H}$ endowed with a special basis. They will be defined in terms of certain graphs. We define a W-graph to be a set of vertices X, with a set Y of edges (an edge is a subset of X consisting of two elements) together with two additional data: for each vertex $x \in X$, we are given a subset I_x of S and, for each ordered pair of vertices y, x such that $\{y, x\} \in Y$, we are given an integer $\mu(y, x) \neq 0$. These data are subject to the requirements (1.0.a), (1.0.b) below. Let E be

* Both authors were supported in part by the National Science Foundation

0020-9910/79/0053/0165/$04.00

the free A-module with basis X. Then

for any $s\in S$

$$\tau_s(x)=\begin{cases}-x, & \text{if } s\in I_x\\ qx+q^{1/2}\sum\limits_{\substack{y\in X\\ s\in I_y\\ \{y,x\}\in Y}}\mu(y,x)\,y, & \text{if } s\notin I_x\end{cases}\tag{1.0.a}$$

defines an endomorphism of E (i.e. the sum over y is assumed to be always finite), and

for any $s\neq t$ in S such that st has finite order m, we require that

$$\underbrace{\tau_s\tau_t\tau_s\ldots}_{m\text{ factors}}=\underbrace{\tau_t\tau_s\tau_t\ldots}_{m\text{ factors}}.\tag{1.0.b}$$

In other words, there is a unique representation $\varphi\colon \mathscr{H}\to\mathrm{End}(E)$ such that $\varphi(T_s)=\tau_s$, for each $s\in S$.

We shall construct, for any W, such a graph. First, we give some definitions. Let $a\to\bar a$ be the involution of the ring A defined by $\overline{q^{1/2}}=q^{-1/2}$. This extends to an involution $h\to\bar h$ of the ring $\mathscr{H}$, defined by

$$\overline{\sum a_w T_w}=\sum \bar a_w T_{w^{-1}}^{-1}.$$

(Note that T_w is an invertible element of $\mathscr{H}$, for any $w\in W$; for example, if $s\in S$, we have $T_s^{-1}=q^{-1}T_s+(q^{-1}-1)$.) For any $w\in W$, we define $q_w=q^{l(w)}$, $\varepsilon_w=(-1)^{l(w)}$. Let $\leqq$ be the usual order relation on W (defined, for example, in [11]). We can now state

Theorem 1.1. *For any $w\in W$, there is a unique element $C_w\in\mathscr{H}$ such that*

$$\bar C_w=C_w\tag{1.1.a}$$

$$C_w=\sum_{y\leqq w}\varepsilon_y\varepsilon_w q_w^{1/2}q_y^{-1}\overline{P_{y,w}}\,T_y\tag{1.1.b}$$

where $P_{y,w}\in A$ is a polynomial in q of degree $\leqq\frac12(l(w)-l(y)-1)$ for $y<w$, and $P_{w,w}=1$.

The following statement is equivalent to Theorem 1.1:

For any $w\in W$, there is a unique element $C'_w\in\mathscr{H}$ such that $\bar C'_w=C'_w$ and $C'_w=q_w^{-1/2}\sum_{y\leqq w}P_{y,w}T_y$, where $P_{y,w}\in A$ is a polynomial in q of degree $\leqq\frac12(l(w)-l(y)-1)$ for $y<w$ and $P_{w,w}=1$. (1.1.c)

The elements C_w and C'_w are related by the identity $C'=\varepsilon_w j(C_w)$, where j is the involution of the ring $\mathscr{H}$ given by $j(\sum a_w T_w)=\sum\bar a_w\varepsilon_w q_w^{-1}T_w$.

It may be conjectured that all coefficients of the polynomial $P_{y,w}$ are non-negative integers.

Definition 1.2 *Given $y,w\in W$ we say that $y\prec w$ if the following conditions are satisfied: $y<w$, $\varepsilon_y=-\varepsilon_w$ and $P_{y,w}$ (given by Theorem 1.1) is a polynomial in q of degree exactly $\frac12(l(w)-l(y)-1)$; in this case, the coefficient of the highest power of q in $P_{y,w}$ is denoted $\mu(y,w)$. It is a non-zero integer. If $w\prec y$, we set $\mu(w,y)=\mu(y,w)$.*

Let W^0 be the group opposed to W. Then $(W \times W^0, S \amalg S^0)$ is a Coxeter group. Let Γ_W be the graph whose vertices are the elements of W and whose edges are the subsets of W of the form $\{y, w\}$ with $y \prec w$. For each $w \in W$, let $I_w = \mathscr{L}(w) \amalg \mathscr{R}(w)^0 \subset S \amalg S^0$, where $\mathscr{L}(w) = \{s \in S \mid sw < w\}$, $\mathscr{R}(w) = \{s \in S \mid ws < w\}$.

Theorem 1.3. *Γ_W, together with the assignment $w \to I_w$ and with the function μ defined above, is a $W \times W^0$-graph.*

Now, given any W-graph Γ, and a subset S' of S, we can regard Γ as W'-graph (where W' is the subgroup of W generated by S') by replacing the set $I_x \subset S$, for each vertex x of Γ, by the set $I_x \cap S'$. In particular, Γ_W can be regarded as a W-graph and as a W^0-graph.

Given any W-graph, Γ, we define a preorder relation $\underset{\Gamma}{\leqq}$ on the set of vertices Γ as follows: we say that the vertices x, x' satisfy $x \underset{\Gamma}{\leqq} x'$, if there exists a sequence of vertices $x = x_0, x_1, \ldots, x_n = x'$ such that for each i, $(1 \leqq i \leqq n)$, $\{x_{i-1}, x_i\}$ is an edge of Γ and $I_{x_{i-1}} \not\subset I_{x_i}$. The equivalence relation on the set of vertices, corresponding to this preorder is denoted $\underset{\Gamma}{\sim}$. (Thus, $x \underset{\Gamma}{\sim} x'$ means that $x \underset{\Gamma}{\leqq} x' \underset{\Gamma}{\leqq} x$.) Each equivalence class, regarded as a full subgraph of Γ (with the same sets I_x and the same function μ) is itself a W-graph. The set of equivalence classes is an ordered set with respect to $\underset{\Gamma}{\leqq}$. In the case of the $W \times W^0$-graph Γ_W, the equivalence classes for $\underset{\Gamma_W}{\sim}$ are called the 2-sided cells of W. When Γ_W is regarded as a W-graph, we shall use the notation $\underset{L}{\leqq}$, $\underset{L}{\sim}$ instead of $\underset{\Gamma_W}{\leqq}$, $\underset{\Gamma_W}{\sim}$; the corresponding equivalence classes are called the left cells of W. When Γ_W is regarded as a W^0-graph, we shall use the notation $\underset{R}{\leqq}$, $\underset{R}{\sim}$ instead of $\underset{\Gamma_W}{\leqq}$, $\underset{\Gamma_W}{\sim}$; the corresponding equivalence classes are called the right cells of W.

In the case where W is the symmetric group s_n, we have

Theorem 1.4. *Let X be a left cell of $W = s_n$, let Γ be the W-graph associated to X and let ρ be the representation of $\mathscr{H}$ (over the quotient field of A) corresponding to Γ. Then ρ is irreducible and the isomorphism class of the W-graph Γ depends only on the isomorphism class of ρ and not on X.*

This gives, in particular, a distinguished basis (defined uniquely up to simultaneous homotety) for any complex irreducible representation of σ_n, with respect to which s_n acts through integral matrices.

Our investigation has started from trying to understand Springer's work connecting unipotent classes and representations of Weyl groups. This had led us to the following question on singularities of Schubert varieties. Let G be a semisimple group over an algebraically closed field, and let $\mathscr{B}$ be the variety of Borel subgroups of G. We fix $B_0 \in \mathscr{B}$, and for each w in the Weyl group W, let $\mathscr{B}_w$ be the set of all $B \in \mathscr{B}$ such that B_0 and B are in relative position w (a Bruhat cell of dimension $l(w)$.) Let $\overline{\mathscr{B}_w}$ be the closure of $\mathscr{B}_w$ (a Schubert variety). Let $T^*(\mathscr{B})$ be the cotangent bundle of $\mathscr{B}$ and let $\mathscr{N}_w \subset T^*(\mathscr{B})$ be the conormal bundle of $\mathscr{B}_w$. Its closure $\overline{\mathscr{N}_w}$ in $T^*(\mathscr{B})$ is an irreducible variety of dimension equal to $\dim(\mathscr{B})$. There is a natural projection $\pi_w: \overline{\mathscr{N}_w} \to \overline{\mathscr{B}_w}$. Now let $y \in W$ be such that $y < w$. Then $\overline{\mathscr{B}_y} \subset \overline{\mathscr{B}_w}$. The question is: for which pairs $y < w$ is it true that $\dim \pi_w^{-1}(\mathscr{B}_y) = \dim(\mathscr{B}) - 1$? It seems likely that when

$G = GL_n$, the condition is precisely that $y \prec w$ in the sense of Definition 1.2. (When $G \neq GL_n$, this is not, in general, true.)

Our polynomials $P_{y,w}$ appear to be very closely connected with the structure of singularities of Schubert varieties. More precisely, $P_{y,w}$ can be regarded as a measure for the failure of local Poincaré duality on the Schubert cell $\overline{\mathscr{B}_w}$ in a neighborhood of a point in $\mathscr{B}_y$. Some results in this direction are formulated in the Appendix.

Another starting point of our investigation was trying to understand the work of Jantzen [6] and Joseph [7, 8] relating primitive ideals in enveloping algebras with representations of Weyl groups.

Let $\underline{g}$ be a semisimple complex Lie algebra. We wish to state a conjecture relating our results with the theory of infinite dimensional representations of $\underline{g}$. We shall need some notations. Let $\underline{h}$ be a Cartan subalgebra of $\underline{g}$ and let $\underline{b}$ be a Borel subalgebra containing $\underline{h}$. Let $\rho: \underline{h} \to \mathbb{C}$ be the linear function on $\underline{h}$ which takes the value 1 on each simple coroot vector. Let W be the Weyl group of $\underline{g}$ with respect to $\underline{h}$ and let S be its set of simple reflections determined by $\underline{b}$. For each $w \in W$, let M_w be the Verma module with highest weight $-w(\rho)-\rho$ and let L_w be its unique irreducible quotient. We can now state

Conjecture 1.5

$$\operatorname{ch} L_w = \sum_{y \leqq w} \varepsilon_y \varepsilon_w P_{y,w}(1) \operatorname{ch} M_y \tag{1.5.a}$$

$$\operatorname{ch} M_w = \sum_{y \leqq w} P_{w_0 w, w_0 y}(1) \operatorname{ch} L_y \tag{1.5.b}$$

for all $w \in W$, where $P_{y,w}$ is the polynomial in q given by Theorem 1.1, *and $P_{y,w}(1)$ denotes its value for $q=1$.*

1.6. *Remarks.* a) The identities (1.5.a) and (1.5.b) are equivalent (see Theorem 3.1).
b) It is known and easy to prove that

$$\operatorname{ch} L_w = \sum_{y \leqq w} \sum_j (-1)^j \dim \operatorname{Ext}^j(M_y, L_w) \operatorname{ch} M_y$$

where Ext is taken in the category $\mathcal{O}$ of Bernstein-Gelfand-Gelfand. (See, for example, [4].) It is also known that $\operatorname{Ext}^j(M_y, L_w) = 0$ if $j > l(w) - l(y)$. (Casselman and Schmid; see also Delorme [4].)

David Vogan has proved [14] that our conjecture 1.5 is equivalent to the formula

$$P_{y,w} = \sum_{i \geqq 0} q^i \dim \operatorname{Ext}^{l(w)-l(y)-2i}(M_y, L_w) \qquad (y \leqq w)$$

and it is also equivalent to the vanishing of $\operatorname{Ext}^j(M_y, L_w)$ for $j \not\equiv l(w) - l(y) \pmod 2$.

c) Conjecture 1.5, together with the results of Joseph [8] and Vogan [13] would imply that the ideal $\operatorname{Ann}(L_w)$ of the universal enveloping algebra of g, annihilating L_w, contains the ideal $\operatorname{Ann}(L_{w'})$ if and only if $w \underset{L}{\leqq} w'$.

1.7. In [6] a distinguished class $\mathcal{S}_W$ of irreducible representations of a Weyl group W was introduced. (Its definition, which will not be reproduced here, was suggested by the representation theory of finite Chevalley groups.) Let X be a left cell of W; it gives rise to a W-graph hence to a representation of $\mathcal{H}$. Specializing $q^{1/2}$ to 1, we get an integral representation of W. The corresponding representation over $\mathbb{Q}$ is not, in general, irreducible. However, it seems likely that it contains a unique irreducible component in the class $\mathcal{S}_W$. We expect that all representations in $\mathcal{S}_W$ are obtained in this way and that two left cells give rise to the same representation in $\mathcal{S}_W$ if and only if they are contained in the same 2-sided cell.

§ 2. The Proofs of Theorems 1.1 and 1.3

Let us define for each $x, y \in W$, an element $R_{x,y} \in A$ by the formula

$$T_{y^{-1}}^{-1} = \sum_x \overline{R_{x,y}}\, q_x^{-1}\, T_x. \tag{2.0.a}$$

The following formulae provide an inductive procedure for computing $R_{x,y}$:

$$R_{x,y} = \begin{cases} R_{sx,sy}, & \text{if } sx<x \text{ and } sy<y \\ R_{xs,ys}, & \text{if } xs<x \text{ and } ys<y \end{cases} \tag{2.0.b}$$

$$R_{x,y} = (q-1)\, R_{sx,y} + q R_{sx,sy}, \quad \text{if } sx>x \text{ and } sy<y. \tag{2.0.c}$$

It follows easily that $R_{x,y} \neq 0$ if and only if $x \leqq y$; when $x \leqq y$, $R_{x,y}$ is a polynomial in q of degree $l(y) - l(x)$. Here are some further properties of $R_{x,y}$.

Lemma 2.1

(i) $\overline{R_{x,y}} = \varepsilon_x \varepsilon_y q_x q_y^{-1} R_{x,y}$.

(ii) $\sum_{x \leqq t \leqq y} \varepsilon_t \varepsilon_x R_{x,t} R_{t,y} = \delta_{x,y}$, *for all* $x \leqq y$ *in* W.

(iii) $R_{x,y} = (q-1)^{l(y)-l(x)}$ *for all* $x \leqq y$ *such that* $l(x) \geqq l(y) - 2$.

(iv) *If* W *is finite and* w_0 *is its longest element, we have* $R_{w_0 y, w_0 x} = R_{x,y}$ *for all* $x, y \in W$.

Proof. (i) follows easily from (2.0.b), (2.0.c). Applying the involution $h \to \bar{h}$ to (2.0.a), we get

$$T_y = \sum_x R_{x,y}\, q_x\, T_{x^{-1}}^{-1}$$

hence the matrices $(R_{x,y}\, q_x)$, $(\overline{R_{x,y}\, q_x^{-1}})$ are inverse to each other. By (i), the matrices $(R_{x,y}\, q_x)$, $(\varepsilon_x \varepsilon_y R_{x,y}\, q_y^{-1})$ are inverse to each other, hence (ii). The formula (iii) is obvious for $x = y$. Assume now that $x \leqq y$ and $l(x) = l(y) - 1$. There is a reduced expression $y = s_1 \dots s_i \dots s_n$ such that $x = s_1 \dots \hat{s}_i \dots s_n$. Using (2.0.b), the computation of $R_{x,y}$ is reduced to the case where $1 = i = n$, in which case (iii) is obvious. Assume now that $x \leqq y$ and $l(x) = l(y) - 2$. There is a reduced expression $y = s_1 \dots s_i \dots s_j \dots s_n$ such that $x = s_1 \dots \hat{s}_i \dots \hat{s}_j \dots s_n$. Using (2.0.b), the computation of $R_{x,y}$ is reduced to the case where $i=1, j=n$. Using then (2.0.c) with $s=s_i$, we see that

$$R_{x,y}=(q-1)\,R_{s_1\ldots\hat{s}_n,\,\hat{s}_1\ldots\hat{s}_n}+qR_{s_1\ldots\hat{s}_n,\,\hat{s}_1\ldots s_n}.$$

By the previous case, $R_{s_1\ldots\hat{s}_n,\,\hat{s}_1\ldots\hat{s}_n}=q-1$; moreover, $R_{s_1\ldots\hat{s}_n,\,\hat{s}_1\ldots s_n}=0$, since $\hat{s}_1\ldots s_n\not\leqq s_1\ldots\hat{s}_n$. Thus, (iii) is proved. (iv) follows easily by induction from (2.0.b), (2.0.c).

2.2. *Proof of Theorem 1.1*

Uniqueness. The equality $\bar{C}_w=C_w$ can be written in the form

$$\sum_{x\leqq w}\varepsilon_x\varepsilon_w q_w^{1/2}q_x^{-1}\bar{P}_{x,w}T_x=\sum_{y\leqq w}\varepsilon_y\varepsilon_w q_w^{-1/2}q_y P_{y,w}\sum_{x\leqq y}\overline{R_{x,y}}\,q_x^{-1}T_x$$

or, equivalently, in the form

$$\varepsilon_x\varepsilon_w q_w^{1/2}q_x^{-1}\bar{P}_{x,w}=\sum_{x\leqq y\leqq w}\varepsilon_y\varepsilon_w q_w^{-1/2}q_y q_x^{-1}\bar{R}_{x,y}P_{y,w}\qquad(\text{for all } x\leqq w).\tag{2.2.a}$$

This is also equivalent to

$$\begin{aligned}&q_w^{1/2}q_x^{-1/2}\bar{P}_{x,w}-q_w^{-1/2}q_x^{1/2}P_{x,w}\\&\quad=\sum_{x<y\leqq w}\varepsilon_x\varepsilon_y q_w^{-1/2}q_y q_x^{-1/2}\bar{R}_{x,y}P_{y,w}\qquad(\text{for all } x<w).\end{aligned}\tag{2.2.b}$$

If the $P_{y,w}$ are known for all y, $x<y\leqq w$ (where $x<w$ is fixed), the Eq. (2.2.b) cannot have more than one solution $P_{x,w}$. Indeed, our assumptions on $P_{x,w}$ imply that $q_w^{-1/2}q_x^{1/2}P_{x,w}$ is a polynomial in $q^{-1/2}$ without constant term, while $q_w^{1/2}q_x^{-1/2}\bar{P}_{x,w}$ is a polynomial in $q^{1/2}$ without constant term. Thus, there cannot be cancellations between these two expressions.

Existence. Clearly, $C_e=T_e$. Assume now that $w\neq e$ and that the existence of C_w satisfying (1.1.a) and (1.1.b) has already been proved for elements w' of length $<l(w)$. We can write $w=sv$, where $s\in S$ and $l(w)=l(v)+1$. Thus C_v is already constructed; the Definition 1.2 can be applied to C_v, so that the relation $z\prec v$ and the corresponding integer $\mu(z,v)$ have a meaning. We now define

$$C_w=(q^{-1/2}T_s-q^{1/2})\,C_v-\sum_{\substack{z\prec v\\ sz<z}}\mu(z,v)\,C_z.$$

To check that C_w satisfies (1.1.a) it is enough to observe that

$$\overline{q^{-1/2}T_s-q^{1/2}}=q^{-1/2}T_s-q^{1/2}.$$

A straightforward computation shows that

$$C_w=\sum_{y\leqq w}\varepsilon_y\varepsilon_w q_w^{1/2}q_y^{-1}\bar{P}_{y,w}T_y$$

where

$$P_{y,w}=q^{1-c}P_{sy,v}+q^c P_{y,v}-\sum_{\substack{z\\ y\leqq z\prec v\\ sz<z}}\mu(z,v)\,q_z^{-1/2}q_v^{1/2}q^{1/2}P_{y,z}\qquad(y\leqq w)\tag{2.2.c}$$

and $c=1$ if $sy<y$, $c=0$ if $sy>y$. (We shall make the convention that $P_{x,v}=0$ when $x \nleq v$.)

(2.2.c) shows that $P_{y,w}$ is a polynomial in q of degree $\leqq \frac{1}{2}(l(v)-l(y))$ if $y<w$ and that $P_{w,w}=1$. Thus, C_w satisfies (1.1.b) and Theorem 1.1 is proved.

2.3. *Proof of Theorem* 1.3. In the process of proving Theorem 1.1, we have seen that

$$T_s C_v = q C_v + q^{1/2} C_{sv} + q^{1/2} \sum_{\substack{z \prec v \\ sz < z}} \mu(z,v) C_z, \qquad \text{if } s\in S \text{ and } sv>v. \tag{2.3.a}$$

A similar proof (interchanging left and right) shows that

$$C_v T_s = q C_v + q^{1/2} C_{vs} + q^{1/2} \sum_{\substack{z \prec v \\ zs < z}} \mu(z,v) C_z, \qquad \text{if } s\in S \text{ and } vs>v. \tag{2.3.b}$$

We now show that

$$T_s C_v = -C_v \qquad \text{if } s\in S \text{ and } sv<v. \tag{2.3.c}$$

We may assume that (2.3.c) is known for elements v' satisfying $sv'<v'$, $l(v')<l(v)$. Using (2.3.a) with v replaced by sv, we see that

$$\begin{aligned} T_s C_v &= T_s\Big(q^{-1/2} T_s C_{sv} - q^{1/2} C_{sv} - \sum_{\substack{z \prec sv \\ sz < z}} \mu(z,sv) C_z\Big) \\ &= q^{-1/2}((q-1)T_s + q) C_{sv} - q^{1/2} T_s C_{sv} + \sum_{\substack{z \prec sv \\ sz < z}} \mu(z,sv) C_z \\ &= q^{1/2} C_{sv} - q^{-1/2} T_s C_{sv} + \sum_{\substack{z \prec sv \\ sz < z}} \mu(z,sv) C_z \\ &= -C_v \end{aligned}$$

as required. An entirely similar proof shows hat

$$C_v T_s = -C_v \qquad \text{if } s\in S \text{ and } vs<v. \tag{2.3.d}$$

To complete the proof of Theorem 1.3 it is now enough to verify the following two statements.

Let $x, y \in W$, $s\in S$ be such that $x<y$, $sy<y$, $sx>x$.
Then $x \prec y$ if and only if $y=sx$.
Moreover, this implies that $\mu(x,y)=1$. (2.3.e)

Let $x, y \in W$, $s\in S$ be such that $x<y$, $ys<y$, $xs<x$.
Then $x \prec y$ if and only if $y=xs$.
Moreover, this implies that $\mu(x,y)=1$. (2.3.f)

Comparing the coefficients of T_{sx} in the two sides of (2.3.c) with $v=y$, we see that,

$$P_{x,y} = P_{sx,y}, \qquad \text{if } x<y,\ sy<y,\ sx>x. \tag{2.3.g}$$

If $sx \neq y$, it follows that $\deg P_{x,y} = \deg P_{sx,y} \leqq \frac{1}{2}(l(y)-l(x)) < \frac{1}{2}(l(y)-l(x)-1)$ hence the relation $x \prec y$ is not satisfied. If $sx=y$, it follows that $P_{x,y}=P_{y,y}=1$, hence $x \prec y$ and

$\mu(x,y)=1$. This proves (2.3.e). The proof of (2.3.f) is entirely similar. This completes the proof of Theorem 1.3.

We now state a property of the preorders $\underset{L}{\leqq}$ and $\underset{R}{\leqq}$ on W.

Proposition 2.4. (i) *If* $x \underset{L}{\leqq} y$, *then* $\mathscr{R}(x) \supset \mathscr{R}(y)$. *Hence, if* $x \underset{L}{\sim} y$, *then* $\mathscr{R}(x)=\mathscr{R}(y)$.

(ii) *If* $x \underset{R}{\leqq} y$, *then* $\mathscr{L}(x) \supset \mathscr{L}(y)$. *Hence, if* $x \underset{R}{\sim} y$, *then* $\mathscr{L}(x)=\mathscr{L}(y)$.

Proof. It is easy to check that, given $s \in S$, we have

$$sy>y \Rightarrow \mathscr{R}(sy) \supset \mathscr{R}(y) \tag{2.4.a}$$

$$ys>y \Rightarrow \mathscr{L}(ys) \supset \mathscr{L}(y). \tag{2.4.b}$$

Assume now that $x \prec y$ and $\mathscr{L}(x) \not\subset \mathscr{L}(y)$. From (2.4.b) we see that $x^{-1}y \notin S$. Using (2.3.f), we see that $\mathscr{R}(x) \supset \mathscr{R}(y)$. This, together with (2.4.a), show that $x \underset{L}{\leqq} y \Rightarrow \mathscr{R}(x) \supset \mathscr{R}(y)$. The proof of (ii) is entirely similar.

2.5. For each $y \leqq w$ in W we define

$$N_{y,w}=q_y \sum_{y \leqq z \leqq w} R_{y,z}. \tag{2.5.a}$$

The following result is stated for future reference.

Lemma 2.6. (i) *For each* $x \leqq y$ *in* W, $P_{x,y}$ *is a polynomial in* q *with constant term* 1.

(ii) *Given* $y<w$ *in* W, *the following two conditions are equivalent:*

$$P_{y',w}=1, \quad \textit{for all } y \leqq y' \leqq w \tag{2.6.a}$$

and

$$N_{y',w}=q_w, \quad \textit{for all } y \leqq y' \leqq w. \tag{2.6.b}$$

(iii) *For each* $y<w$ *such that* $l(w)=l(y)+1$, *we have* $N_{y,w}=q_w$ *and* $P_{y,w}=1$. *In particular, we have* $y \prec w$ *and* $\mu(y,w)=1$.

(iv) *For each* $y<w$ *such that* $l(w)=l(y)+2$, *we have* $N_{y,w}=q_w$ *and* $P_{y,w}=1$.

(v) *For each* $w \in W$, *we have*

$$q_w^{-1} \sum_{y \leqq w} q_y P_{y,w} = \overline{\sum_{y \leqq w} q_y P_{y,w}}.$$

(vi) *If* W *is finite and* w_0 *is its longest element, then* $P_{y,w_0}=1$ *for all* $y \in W$.

Proof. (i) follows immediately from the inductive formula (2.2.c). To prove (ii), we may assume, by induction on $l(w)-l(y)$, that $P_{y',w}=1$ for all y' such that $y<y' \leqq w$. Then, the identity

$$P_{y,w}=\sum_{y \leqq y' \leqq w} \varepsilon_y \varepsilon_{y'} R_{y,y'} \overline{P_{y',w}}\, q_{y'}^{-1} q_w$$

(see 2.2.a) becomes

$$P_{y,w}=q_y^{-1}q_w\overline{P_{y,w}}+\sum_{y<y'\leqq w}\varepsilon_y\varepsilon_{y'}R_{y,y'}q_{y'}^{-1}q_w.$$

Using Lemma 2.1(i) this can be also written as

$$\begin{aligned}P_{y,w}&=q_y^{-1}q_w\overline{P_{y,w}}+\sum_{y<y'\leqq w}\overline{R_{y,y'}q_y^{-1}}q_w\\&=q_y^{-1}q_w\overline{P_{y,w}}-q_y^{-1}q_w+q_w\sum_{y\leqq y'\leqq w}\overline{R_{y,y'}q_y},\end{aligned}$$

hence

$$P_{y,w}-q_y^{-1}q_w\overline{P_{y,w}}=q_w\overline{N_{y,w}}-q_y^{-1}q_w. \tag{2.6.c}$$

If $P_{y,w}=1$, it follows that $N_{y,w}=q_w$. Conversely, if $N_{y,w}=q_w$, it follows that

$$q_y^{1/2}q_w^{-1/2}(P_{y,w}-1)=\overline{q_y^{1/2}q_w^{-1/2}(P_{y,w}-1)}.$$

But $q_y^{1/2}q_w^{-1/2}(P_{y,w}-1)$ is a polynomial in $q^{-1/2}$ without constant term; therefore it can be fixed by the involution $a\to\bar{a}$ only if it is zero. It follows that $P_{y,w}=1$ and (ii) is proved. Using Lemma 2.1(iii), we see that with the assumptions of (iii) we have $N_{y,w}=q_yR_{y,y}+q_yR_{y,w}=q_y+q_y(q-1)=q_w$. Using (ii), we deduce that $P_{y,w}=1$, hence (iii). Under the assumptions of (iv), it is known that there are exactly two elements z_1, z_2 such that $y<z_1<w$, $y<z_2<w$. Using Lemma 2.1(iii), we see that

$$N_{y,w}=q_y(R_{y,y}+R_{y,z_1}+R_{y,z_2}+R_{y,w})=q_y(1+(q-1)+(q-1)+(q-1)^2)=q_w.$$

Using (ii) and the fact that $N_{z_1,w}=N_{z_2,w}=q_w$ (given by (iii)) it follows that $P_{y,w}=1$. The identity (v) is just the identity $\mathscr{X}(C_w)=\mathscr{X}(\bar{C}_w)$, where $\mathscr{X}\colon \mathscr{H}\to A$ is the algebra homomorphism defined by $\mathscr{X}(T_y)=\varepsilon_y$ for all y. (vi) follows by applying repeatedly (2.3.g).

§ 3. An Inversion Formula

Our next result describes, in the case where W is finite, the inverse of the triangular matrix $(P_{x,y})$, where $P_{x,y}$ is defined to be zero if $x\nleqq y$.

Theorem 3.1. *Assume that W is finite and let w_0 be its longest element. We have*

$$\sum_{x\leqq z\leqq y}\varepsilon_x\varepsilon_z P_{x,z}P_{w_0y,w_0z}=\delta_{x,y}, \quad \textit{for all } x\leqq y \textit{ in } W. \tag{3.1.a}$$

Proof. Let $M_{x,y}$ be the left hand side of (3.1.a). We may assume that $x<y$ and that $M_{t,s}=0$ for all $t<s$ such that $l(s)-l(t)<l(y)-l(x)$. We start with the identity (2.2.a):

$$P_{x,z}=\sum_{x\leqq t\leqq z}\varepsilon_x\varepsilon_t R_{x,t}\overline{P_{t,z}}\,q_t^{-1}q_z \quad (x\leqq z \text{ in } W).$$

It follows that

$$M_{x,y}=\sum_{x\leqq z\leqq y}\varepsilon_x\varepsilon_z\sum_{\substack{x\leqq t\leqq z\\ z\leqq s\leqq y}}\varepsilon_x\varepsilon_t\varepsilon_y\varepsilon_s R_{x,t}\overline{P_{t,z}}R_{w_0y,w_0s}\overline{P_{w_0s,w_0z}}q_t^{-1}q_zq_z^{-1}q_s$$
$$=\sum_{\substack{t,s\\ x\leqq t\leqq s\leqq y}}\varepsilon_y\varepsilon_s q_t^{-1}q_sR_{x,t}R_{w_0y,w_0s}\overline{M_{t,s}}.$$

The only t,s which can contribute to this sum satisfy $t=s$ or $t=x$, $s=y$. Thus,

$$M_{x,y}=q_x^{-1}q_y\overline{M_{x,y}}+\sum_{x\leqq t\leqq y}\varepsilon_y\varepsilon_tR_{x,t}R_{w_0y,w_0t}.$$

Using Lemma 2.1(iv) and (ii), we see that the last sum (over t) equals

$$\sum_{x\leqq t\leqq y}\varepsilon_y\varepsilon_tR_{x,t}R_{t,y}=0.$$

Thus $M_{x,y}=q_x^{-1}q_y\overline{M_{x,y}}$ hence $q_x^{1/2}q_y^{-1/2}M_{x,y}=q_x^{-1/2}q_y^{1/2}\overline{M_{x,y}}$. The bounds on the degree of the polynomials $P_{y,w}$ described in Theorem 1.1 imply that $\overline{q_x^{-1/2}q_y^{1/2}M_{x,y}}$ is a polynomial in $q^{1/2}$ without constant term. Hence it cannot be fixed by the involution $a\to\bar a$, unless it is zero. Thus, $M_{x,y}=0$, as required.

Corollary 3.2. *Let $x<y$ be two elements of W (assumed to be finite). The following conditions are equivalent: $x\prec y$ and $w_0y\prec w_0x$. If these conditions are satisfied, we have $\mu(x,y)=\mu(w_0y,w_0x)$.*

Proof. We can assume that $\varepsilon_x=-\varepsilon_y$. The difference $P_{w_0y,w_0x}-P_{x,y}$ is equal to $\sum_{x<z<y}\varepsilon_x\varepsilon_zP_{x,z}P_{w_0y,w_0z}$ and one checks easily that the last expression is a polynomial in q of degree $<\frac12(l(y)-l(x)-1)$. Therefore, the $\frac12(l(y)-l(x)-1)$-th power of q appears in $P_{x,y}$ with the same coefficients as in P_{w_0y,w_0x}.

3.3. *Remarks.* a) The map $x\to w_0x$ reverses each of the preorders $\underset{L}{\leqq},\underset{R}{\leqq},\underset{LR}{\leqq}$ on W. Hence it induces an order reversing involution on the set of left cells of W, on the set of right cells of W and on the set of 2-sided cells of W.

b) Setting $q=0$ in the identity (3.1.a) and using Lemma 2.6(i) we get the following known identity [11]:

$$\sum_{x\leqq z\leqq y}\varepsilon_x\varepsilon_z=\delta_{x,y},\qquad\text{for all } x\leqq y \text{ in } W.$$

§ 4. Some Preliminaries to the Proof of Theorem 1.4

4.1. Let us fix two reflections s,t in S such that st has order 3. Let

$\mathscr{D}_L(s,t)=\{w\in W\mid\mathscr{L}(w)\cap\{s,t\}$ has exactly one element$\}$

$\mathscr{D}_R(s,t)=\{w\in W\mid\mathscr{R}(w)\cap\{s,t\}$ has exactly one element$\}$.

If $w\in\mathscr{D}_L(s,t)$, then exactly one of the elements sw,tw is in $\mathscr{D}_L(s,t)$; we denote it $^*w$. The map $w\to{}^*w$ is an involution of $\mathscr{D}_L(s,t)$. Similarly, we have an involution

$w \to w^*$ of $\mathscr{D}_R(s,t)$: w^* is the unique element of $\mathscr{D}_R(s,t) \cap \{ws, wt\}$. Let $\langle s,t \rangle$ be the group of order 6 generated by s,t. We shall prove

Theorem 4.2. *Let y, w be two elements in $\mathscr{D}_L(s,t)$.*

(i) *If $yw^{-1} \notin \langle s,t \rangle$, then we have $y \prec w$ if and only if $^*y \prec {}^*w$, and then $\mu(y,w) = \mu(^*y, {}^*w)$.*

(ii) *If $yw^{-1} \in \langle s,t \rangle$, then we have $y \prec w$ if and only if $^*w \prec {}^*y$, and then $\mu(y,w) = \mu(^*w, {}^*y) = 1$.*

Let y, w be two elements in $\mathscr{D}_R(s,t)$.

(iii) *If $y^{-1}w \notin \langle s,t \rangle$, then we have $y \prec w$ if and only if $y^* \prec w^*$, and then $\mu(y,w) = \mu(y^*, w^*)$.*

(iv) *If $y^{-1}w \in \langle s,t \rangle$, then we have $y \prec w$ if and only if $w^* \prec y^*$, and then $\mu(y,w) = \mu(w^*, y^*) = 1$.*

Proof. Throughout this proof, we shall use the following notations. For any $x < x'$ in W such that $\varepsilon_x = -\varepsilon_{x'}$, we set $d(x,x') = \frac{1}{2}(l(x') - l(x) - 1)$ and let $\mu(x,x')$ be the coefficient of $g^{d(x,x')}$ in $P_{x,x'}$. Thus $x \prec x'$ if and only if $\mu(x,x') \neq 0$. If P' is a polynomial in q, we say that $P_{x,x'} \sim P'$ if $P_{x,x'} - P'$ is of degree $< d(x,x')$. In particular, $P_{x,x'} \sim \mu(x,x')\, q^{d(x,x')}$.

It is enough to prove statements (i) and (ii). With the assumptions of (ii), we have $y \prec w$ if and only if $y < w$ and $l(w) = l(y) + 1$ and then $\mu(y,w) = 1$. (See Lemma 2.6(iii).) The conclusion of (ii) follows immediately. In the remainder of the proof we shall assume that $y, w \in \mathscr{D}_L(s,t)$ and $yw^{-1} \notin \langle s,t \rangle$. We may assume that $\varepsilon_y = -\varepsilon_w$. (This is equivalent to $\varepsilon_{^*y} = -\varepsilon_{^*w}$.) There are two cases to consider.

Case 1. $\quad {}^*y \cdot y^{-1} = {}^*w \cdot w^{-1}$.

In this case, we may assume without loss of generality that $tsy < sy < y < ty$ and $tsw < sw < w < tw$, so that $^*y = sy$, $^*w = sw$. It is clear that the conditions $y < w$ and $sy < sw$ are equivalent. Thus, we may assume that $y < w$. From (2.2.c), it follows that $P_{y,w} = P_{sy,sw}$ if $y \nleq sw$ and

$$P_{y,w} \sim P_{sy,sw} + q P_{y,sw} - \sum_{\substack{y \prec z \prec sw \\ sz < z}} \mu(y,z)\,\mu(z,sw)\, q^{d(y,w)} \tag{4.2.a}$$

if $y \leqq sw$. Thus, we can assume that $y \leqq sw$. This implies that $ty \leqq sw$, since $t \in \mathscr{L}(sw)$. From (2.3.e), we see that for any z in the last sum, such that $z \neq ty$, $z \neq tws$, we have $t \in \mathscr{L}(sw) \Rightarrow t \in \mathscr{L}(z) \Rightarrow t \in \mathscr{L}(y)$, a contradiction. On the other hand, $z = ty$ satisfies $sz < z$, while $z = tws$ doesn't satisfy $sz < z$. Thus the sum over z has exactly one term: $z = ty$. We have $\mu(y, ty) = 1$, hence (4.2.a) becomes

$$P_{y,w} \sim P_{sy,sw} + q P_{y,sw} - \mu(ty, sw)\, q^{d(y,w)}.$$

By (2.3.g), we have $P_{y,sw} = P_{ty,sw}$ so that $q P_{y,sw} - \mu(ty,sw)\, q^{d(y,w)}$ is a polynomial in q of degree $< d(y,w)$. It follows that $P_{y,w} \sim P_{sy,sw}$ as required.

Case 2. $\quad {}^*y \cdot y^{-1} \neq {}^*w \cdot w^{-1}$.

In this case, we may assume without loss of generality that $tsy < sy < y < ty$, $sw < w < tw < stw$, so that $^*y = sy$, $^*w = tw$. We can clearly assume that $sy < tw$:

This implies that $tsy<w$ and $y<stw$. From (2.2.c), it follows that

$$P_{sy,tw}=P_{tsy,w} \quad \text{if } sy\nleqq w$$

and

$$P_{sy,tw}\sim P_{tsy,w}+qP_{sy,w}-\sum_{\substack{sy\prec z\prec w\\ tz<z}}\mu(sy,z)\,\mu(z,w)\,q^{d(sy,tw)}$$

if $sy\leqq w$. We have $s\in\mathscr{L}(w)$, $s\notin\mathscr{L}(tsy)$ and $w\neq stsy$, hence, by (2.3.e), the relation $tsy\prec w$ cannot hold. Thus, if $sy\nleqq w$, we have $P_{sy,tw}\sim 0$, hence $sy\prec tw$ fails to be true. On the other hand, if $sy\nleqq w$, we must have also $y\nleqq w$ (since $s\in\mathscr{L}(w)$), hence $y\prec w$ also fails to be true. Thus, we may assume that $sy\leqq w$, so that

$$P_{sy,tw}\sim qP_{sy,w}-\sum_{\substack{sy\prec z\prec w\\ tz<z}}\mu(sy,z)\,\mu(z,w)\,q^{d(sy,tw)}.$$

From (2.3.e) we see that for any z in the last sum, such that $z\neq y$, $z\neq sw$, we have $s\in\mathscr{L}(w)\Rightarrow s\in\mathscr{L}(z)\Rightarrow s\in\mathscr{L}(sy)$, a contradiction. On the other hand, neither $z=y$ nor $z=sw$ satisfy $tz<z$. It follows that $P_{sy,tw}\sim qP_{sy,w}$. By (2.3.g), we have $P_{sy,w}=P_{y,w}$ (we must have $y\leqq w$, since $sy\leqq w$ and $s\in\mathscr{L}(w)$). Thus, $P_{sy,tw}\sim qP_{y,w}$, hence $\mu(sy,tw)=\mu(y,w)$, as required.

Corollary 4.3. (i) *Let y, w be two elements in $\mathscr{D}_L(s,t)$. If $y\underset{R}{\sim}w$, then $^*y\underset{R}{\sim}w^*$.*

(ii) *Let y, w be two elements in $\mathscr{D}_R(s,t)$. If $y\underset{L}{\sim}w$, then $y^*\underset{L}{\sim}w^*$.*

Proof. We first note that, if $x\in\mathscr{D}_L(s,t)$, then $^*x\underset{L}{\sim}x$, hence, by Proposition 2.4(i), we have $\mathscr{R}(^*x)=\mathscr{R}(x)$. Now let y, w be two elements in $\mathscr{D}_L(s,t)$ such that $y\underset{R}{\sim}w$. Then there exists a sequence $y=y_1,y_2,\ldots,y_n=w$ such that $\{y_i,y_{i+1}\}$ is an edge of Γ_W and $\mathscr{R}(y_i)\not\subset\mathscr{R}(y_{i+1})$ for $i=1,\ldots,n-1$, and there exists a sequence $w=w_1,w_2,\ldots,w_m=y$ such that $\{w_j,w_{j+1}\}$ is an edge of Γ_W and $\mathscr{R}(w_j)\not\subset\mathscr{R}(w_{j+1})$ for $j=1,\ldots,m-1$. Clearly, all elements y_i, w_j are in the same right cell, hence, by Proposition 2.4(ii), we have $\mathscr{L}(y_i)=\mathscr{L}(y)$ for all i, $\mathscr{L}(w_j)=\mathscr{L}(y)$ for all j. Since $y\in\mathscr{D}_L(s,t)$, it follows that $y_i\in\mathscr{D}_L(s,t)$ for all i and $w_j\in\mathscr{D}_L(s,t)$ for all j. Hence $^*y_i$ and $^*w_j$ are well defined. Theorem 4.2 shows that $\{^*y_i,{}^*y_{i+1}\}$ is an edge of Γ_W for $i=1,\ldots,n-1$ and that $\{^*w_j,{}^*w_{j+1}\}$ is an edge of Γ_W for $j=1,\ldots,m-1$. By the remark at the beginning of the proof, we have $\mathscr{R}(y_i)=\mathscr{R}(^*y_i)$ for all i and $\mathscr{R}(w_j)=\mathscr{R}(^*w_j)$ for all j. It follows that $\mathscr{R}(^*y_i)\not\subset\mathscr{R}(^*y_{i+1})$ for $i=1,\ldots,n-1$ and $\mathscr{R}^*(w_j)\not\subset\mathscr{R}(^*w_{j+1})$ for $j=1,\ldots,m-1$. This shows that $^*y={}^*y_1\underset{R}{\leqq}{}^*y_2\underset{R}{\leqq}\cdots\underset{R}{\leqq}{}^*y_n=w={}^*w_1\underset{R}{\leqq}{}^*w_2\underset{R}{\leqq}\cdots\underset{R}{\leqq}{}^*w_m={}^*y$ hence $^*y\underset{R}{\sim}{}^*w$ and (i) is proved. The proof of (ii) is entirely similar.

§ 5. Proof of Theorem 1.4

In [12], Vogan defines for any Weyl group W, an equivalence relation on W by means of a "generalized τ-invariant". In his language, Corollary 4.3 can be reformulated to say that two elements $y, w \in W$ such that $y \underset{L}{\sim} w$, must have the same generalized τ-invariant (provided that the Coxeter graph of W is simply laced). Moreover, in the case where W is the symmetric group s_n, Jantzen and Vogan have shown [loc. cit., Thm. 6.5] that if $y, w \in W$ have the same generalized τ-invariant, then $y \approx w$, where $\approx$ is the equivalence relation generated by the relations $x \approx sx$ where $s \in S$, $x < sx$, $\mathscr{L}(x) \not\subset \mathscr{L}(sx)$. On the other hand, it is clear that two elements equivalent under $\approx$ are equivalent under $\underset{L}{\sim}$. Thus, for $W = s_n$, the equivalence relations $\underset{L}{\sim}$ and $\approx$ coincide. The equivalence relation $\approx$ on s_n has been studied by combinatorists (see, for example [9, 5.1.4 and Ex. 5]). The following result is known: If X is an equivalence class for $\approx$ (i.e. a left cell) and if y, y' are distinct elements of X^{-1}, then the $\approx$ equivalence classes $X_y, X_{y'}$ containing y, y' respectively, are disjoint; moreover $X = X_y$ for some $y \in X^{-1}$. We now show that the W-graphs $\Gamma_y, \Gamma_{y'}$ associated to the left cells $X_y, X_{y'}$ $(y, y' \in X^{-1})$ are isomorphic. We have $y^{-1} \approx y'^{-1}$, hence, by the definition of $\approx$, we are reduced to the case where there exist $s, t \in S$ such that $(st)^3 = 1$, $y \in \mathscr{D}_R(s, t)$ and $y^* = y'$ ($*$ defined with respect to s, t). It follows that all elements of X_y and of $X_{y'}$ are in $\mathscr{D}_R(s, t)$ (cf. Proposition 2.4(i)) and that $w \to w^*$ is a bijection of X_y onto $X_{y'}$ (cf. Corollary 4.3(ii)). It defines an isomorphism between the W-graphs $\Gamma_y, \Gamma_{y'}$ (cf. Theorem 4.2). In particular, for any $y, y' \in X^{-1}$, the representations $\rho_y, \rho_{y'}$ of $\mathscr{H}$ associated to $\Gamma_y, \Gamma_{y'}$ are isomorphic. The sum of the representations of $\mathscr{H}$ associated to the various left cells is equal to the regular representation (over some field containing A). If ρ is the representation corresponding to X, then $\sum_{y \in X^{-1}} \rho_y = (\dim \rho)\, \rho$ is a subrepresentation of the regular representation. It follows that ρ is irreducible, and that the left cells which give rise to a representation isomorphic to ρ are exactly the left cells X_y $(y \in X^{-1})$. This completes the proof of Theorem 1.4.

§ 6. Examples

6.1. Let W be a Weyl group of type A_3 with $S = \{s_1, s_2, s_3\}$ such that $s_1 s_3 = s_3 s_1$. There are exactly two pairs of elements $y < w$ in W such that $y \prec w$, $l(w) - l(y) > 1$. These are $s_2 < s_2 s_1 s_3 s_2$ and $s_1 s_3 < s_1 s_3 s_2 s_3 s_1$. For both pairs we have $P_{y,w} = 1 + q$.

6.2. Let (W, S) be a Coxeter group such that for any $s \neq t$ in S, the order $m_{s,t}$ of st is $2, 3, 4, 6$ or ∞. There is a standard graph Γ associated to (W, S): its set of vertices is S and $\{s, t\}$ is an edge precisely when $m_{s,t} \geqq 3$. We associate to each $s \in S$ the set $I_s = \{s\}$ and we consider a function μ on the set of ordered pairs, s, t which are joined in Γ such that $\mu(s, t)\, \mu(t, s) = 4 \cos^2 \pi/m_{s,t}$. This is a W-graph (see the work of Kilmoyer [3]).

We shall now give some examples of W-graphs associated to left cells in the Coxeter group (W, S). In all these examples, the function μ is identically 1, hence it will be omitted. The vertices will be represented by circles, inside which we describe the corresponding subset of S.

If W is of type A_2 with Coxeter graph $\overset{1}{\vdash}\!\!\text{——}\!\!\overset{2}{\dashv}$, the W-graphs arising from the left cells of W are:

$\ominus$, ①——②, (1, 2).

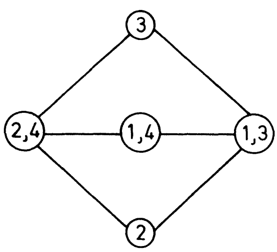
If W is of type B_2 with Coxeter graph $\overset{1}{\vdash}\!\!\text{══}\!\!\overset{2}{\dashv}$, the W-graphs arising from the left cells of W are:

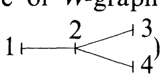
$\ominus$, ①——②——①, ②——①——②, (1, 2).

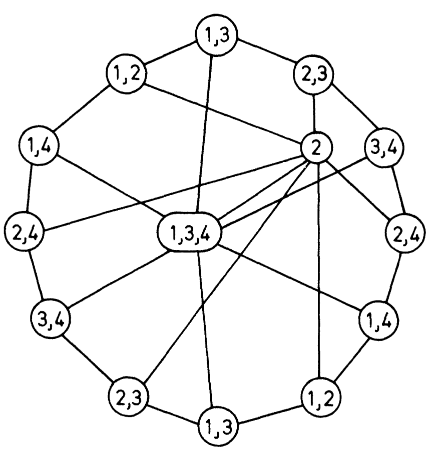
If W is of type A_3, with Coxeter graph $\overset{1}{\vdash}\!\!\text{——}\!\!\overset{2}{+}\!\!\text{——}\!\!\overset{3}{\dashv}$, the W-graphs arising from the left cells of W are:

$\ominus$, ①——②——③, (1, 2)——③, (2, 3)——(1, 3)——(1, 2), (1, 2, 3).

An example of W-graph associated to a left cell of W of type A_4 (with Coxeter graph $\overset{1}{\vdash}\!\!\text{——}\!\!\overset{2}{+}\!\!\text{——}\!\!\overset{3}{+}\!\!\text{——}\!\!\overset{4}{\dashv}$):

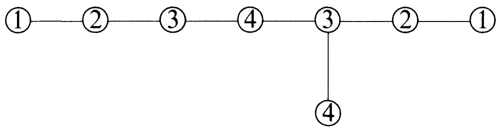


An example of W-graph associated to a left cell of W of type D_4 (with Coxeter graph ):

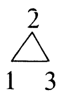


An example of W-graph associated to a left cell of the non-crystalographic finite Coxeter group W of type H_4 with simple reflections s_1, s_2, s_3, s_4 such that

$(s_1 s_2)^3 = (s_2 s_3)^3 = (s_3 s_4) \ = 1$:



(6.2.c)

An example of W-graph associated to a left cell of the affine Weyl group W with Coxeter graph  :

(6.2.d)

6.3. Let u be a unipotent element in $GL_n(\mathbb{C})$ and let $\mathscr{B}_u$ be the variety of Borel subgroups containing u. Let X be the set of irreducible components of $\mathscr{B}_u$. We associate to u a graph Γ_u as follows: the set of vertices of Γ_u is X; two vertices are joined precisely when the corresponding components of $\mathscr{B}_u$ have an intersection of dimension equal to $\dim(\mathscr{B}_u) - 1$. To each component C of $\mathscr{B}_u$, we associate a set I_C of simple reflections in the Weyl group W as follows. We identify the set of simple reflections in W with the set of conjugacy classes of rank 1 parabolic subgroups. Let $\mathscr{P}_s$ be the class of parabolic subgroups corresponding to s and let π_s: $\mathscr{B} \to \mathscr{P}_s$ be the natural projection. We say that $s \in I_C$, if C is a union of fibres of the map π_s. We have verified that, for $n \leqq 6$, Γ_u, together with the assignment $C \to I_C$ and with the function $\mu \equiv 1$ is a W-graph, and that the W-graphs obtained in this way from the various unipotent classes in $GL_n(\mathbb{C})$ are the same as the W-graphs associated to the left cells of W. We have also shown that the graph (6.2.b) has an analogous geometric interpretation in terms of a unipotent class in $SO_8(\mathbb{C})$.

Appendix

We will discuss here some algebraic geometry related to the polynomials $R_{y,w}$ and $P_{y,w}$.

The lemmas in this Appendix are not difficult to prove and their proofs will be generally omitted.

Let k be an algebraic closure of the prime field F_p with p elements. We will consider algebraic varieties over k. For any such variety we denote by $H^*(X)$ the étale cohomology of X with values in the constant sheaf $\mathbb{Q}_l$, where l is a fixed prime $\neq p$. If $x \in X$, we denote by $H^*_{\langle x \rangle}(X)$ the cohomology of X with support in x (see [5, Éxp. 13, p. 2]). There is an exact sequence

$$\ldots \to H^i(X) \to H^i(X - x) \to H^{i+1}_{\langle x \rangle}(X) \to H^{i+1}(X) \to \ldots$$

If X is non-singular, of dimension d at x, then $H^i_{\langle x \rangle}(X) = 0$ if $i \neq 2d$ and $H^{2d}_{\langle x \rangle}(X) = \mathbb{Q}_l(-d)$.

Definition A1. *Let X be an irreducible variety of dimension d.*

(a) *X is rationally smooth if for all $x \in X$, we have $H^i_{\langle x\rangle}(X)=0$ if $i \neq 2d$ and $H^{2d}_{\langle x\rangle}(X)=\mathbb{Q}_l(-d)$.*

(b) *We say that $x \in X$ is a rationally smooth point of X if there exists an open, rationally smooth neighborhood of x in X.*

(c) *We denote by $\mathcal{S}(X) \subset X$ the set of all points of X which are not rationally smooth.*

$\mathcal{S}(X)$ is a closed subset of X, contained in the set $\mathrm{Sing}(X)$ of singular points of X.

Let G be a semisimple adjoint algebraic group over k. Let B_0 be a Borel subgroup, $T_0 \subset B_0$ a maximal torus and let W be the corresponding Weyl group.

The set $\mathcal{B}$ of Borel subgroups of G has a natural structure of projective G-variety: $(g,B) \to B^g$. The set $\mathcal{B}^{T_0} \subset \mathcal{B}$ of T_0-invariant points is in 1-1 correspondence with W: $w \leftrightarrow B_0^w$. Given two points B_1, B_2 in $\mathcal{B}$, we say that B_1, B_2 are in relative position w ($w \in W$) if, for some $g \in G$, we have $B_1^g=B_0$, $B_2^g=B_0^w$ (we then write $B_1 \xrightarrow{w} B_2$). For any $w \in W$, we denote by $\mathcal{B}_w$ the set of all $B \in \mathcal{B}$ such that $B_0 \xrightarrow{w} B$. Its closure $\overline{\mathcal{B}_w}$ is called a Schubert variety. It is known that, given two elements $y, w \in W$, we have $y \leqq w$ if and only if $\overline{\mathcal{B}_y} \subset \overline{\mathcal{B}_w}$

Theorem A2. *Given $y<w$ in W, the following conditions are equivalent:*

(a) $\mathcal{B}_y \cap \mathcal{S}(\overline{\mathcal{B}_w}) = \emptyset$

(b) $N_{y',w}=q_w$, *for all y', $y \leqq y' \leqq w$*

(c) $P_{y',w}=1$, *for all y', $y \leqq y' \leqq w$.*

We have seen already (Lemma 2.6(ii)) that (b), (c) are equivalent. By induction on $l(w)-l(y)$, we can restrict ourselves to the case where, for all y', $y<y' \leqq w$, we have $\mathcal{B}_{y'} \cap \mathcal{S}(\overline{\mathcal{B}_w})=\emptyset$ and $P_{y',w}=1$. In the rest of the proof, y and w are fixed.

We now fix an F_p-rational structure on G such that G is F_p-split and T_0, B_0 are defined over F_p. Then $\mathcal{B}, \mathcal{B}_W$ and, more generally, all algebraic varieties X we will deal with will be F_p-varieties. For such a variety, we denote by $|X|_r$ the number of F_{p^r}-rational points of X.

The Hecke algebra $\mathcal{H}$ will enter in the proof by means of the following Lemma. For any tripe $w_1, w_2, w_3 \in W$, let $\mathcal{N}(w_1, w_2, w_3) \subset \mathcal{B}$ be the set of all points $B \in \mathcal{B}$ such that $B_0 \xrightarrow{w_2} B \xrightarrow{w_3} B_0^{w_1}$.

Lemma A3. *There is a unique polynomial $c(W_1, w_2, w_3; q)$ such that $c(w_1, w_2, w_3; p^r)=|\mathcal{N}(w_1, w_2, w_3)|_r$ for all $r \geqq 1$. We have*

$$T_{w_2} T_{w_3} = \sum_{w_1 \in W} c(w_1, w_2, w_3; q)\, T_{w_1}.$$

For any $w', w'' \in W$, we denote by $\mathcal{B}_{w'}(w'')$ the set of points $B \in \mathcal{B}$ such that $B_0^{w''} \xrightarrow{w'} B$. We define

$$U=\overline{\mathcal{B}_w} \cap \mathcal{B}_{w_0}(y w_0), \qquad V=\overline{\mathcal{B}_w} \cap \mathcal{B}_{w_0 y}(w_0).$$

Then U is an open neighborhood of $\mathcal{B}_y$ in $\overline{\mathcal{B}_w}$. We have $U^{T_0}=V^{T_0}=\{b\}$, where $b = B_0^y$.

Lemma A4. (a) $|U|_r = N_{y,w}(p^r)$.

(b) *There is a canonical T_0-invariant isomorphism*

$$U \simeq \mathscr{B}_y \times V.$$

Part (a) follows from Lemma A3 and the following elementary statement: the polynomials $R_{w',w}$ (see (2.0.a)) satisfy the identity

$$T_w T_{w_0} = \sum_{w' \leqq w} R_{w',w} q_{w'} T_{w' w_0}.$$

The isomorphism in (b) is the restriction of an isomorphism

$$\mathscr{B}_{w_0}(y w_0) \cong \mathscr{B}_y \times \mathscr{B}_{w_0 y}(w_0).$$

Remark A5. Our assumptions imply that $V-0$ is rationally smooth.

Lemma A6. *There exists an F_p-isomorphism of algebraic varieties φ: $L \simeq k^n$ (where $n = l(w_0 y)$) such that*

(a) $\varphi(b) = 0$,

(b) *the induced action of T_0 on k^n is given by*

$$t: (e_1, \ldots, e_n) \to (\chi_1(t) e_1, \ldots, \chi_n(t) e_n)$$

where $\chi_1, \ldots, \chi_n$ are characters of T_0.

(c) *There exists an imbedding j: $\mathbb{G}_m \to T_0$ such that for all i, $1 \leqq i \leqq n$, the composition $\chi_i \cdot j$: $\mathbb{G}_m \to \mathbb{G}_m$ is given by $\lambda \to \lambda^{a_i}$, $a_i > 0$.*

We will identify L with k^n via φ and V with the corresponding subvariety of k^n. We will regard L and V as G_m-varieties.

Lemma A7. *Let Z be an algebraic variety with an action ψ: $\mathbb{G}_m \times Z \to Z$ of $\mathbb{G}_m$ and let $z_0 \in Z$ be a $\mathbb{G}_m$-invariant point. Suppose that $\mathbb{G}_m$ "contracts Z to z_0" i.e. ψ can be extended to a morphism*

$$\begin{array}{ccc} \tilde{\psi}\colon \mathbb{A}^1 \times Z & \longrightarrow & Z \\ \uparrow & & \| \\ \psi\colon \mathbb{G}_m \times Z & \longrightarrow & Z \end{array}$$

such that $\tilde{\psi}(0 \times Z) = z_0$. Then $H^i_{\langle z_0 \rangle}(Z) \cong H^{i-1}(Z - z_0)$ for $i \neq 1$ and $H^1_{\langle z_0 \rangle}(Z) = 0$.

Definition A8. (a) *Let Y be an affine algebraic variety with a $\mathbb{G}_m$-action. We say that this action is standard if there exists a finite group $\Gamma \subset \mathbb{G}_m$, a variety Y_0 and an action of Γ on Y_0 such that Y is isomorphic as a $\mathbb{G}_m$-variety to $\Gamma \backslash (\mathbb{G}_m \times Y_0)$ where Γ acts diagonally on $\mathbb{G}_m \times Y_0$.*

(b) *We say that an action of $\mathbb{G}_m$ on an algebraic variety X is locally standard if there is a covering $X = Y_1 \cup Y_2 \cup \ldots \cup Y_m$, where Y_i are open, $\mathbb{G}_m$-invariant affine subsets of X such that the action of $\mathbb{G}_m$ on each Y_i is standard.*

Lemma A9. *Let X be an algebraic variety with a locally standard action of $\mathbb{G}_m$. Then*

(a) *the geometric quotient $\pi\colon X \to \hat{X}$ (=set of $\mathbb{G}_m$-orbits on X) exists.*

(b) *$R^i\pi_*(\mathbb{Q}_l)$ is zero if $i \neq 0, 1$ and is isomorphic to $\mathbb{Q}_l$ (resp. $\mathbb{Q}_l(-1)$) for $i=0$ (resp. $i=1$).*

(c) *X is rationally smooth if and only if $\hat{X}$ is rationally smooth.*

(d) *$|X|_r = (p^r - 1)|\hat{X}|_r$, for all $r \geqq 1$.*

(e) *If X' is a $\mathbb{G}_m$-invariant closed subset of X, then the $\mathbb{G}_m$ action on X' is locally standard.*

Lemma A10. *The action of $\mathbb{G}_m$ on $L-0$ (and, hence on $V-0$) is locally standard.*

Consider the geometric quotients (for the $\mathbb{G}_m$-action) $\pi\colon L-0 \to \hat{L}$ and $\pi\colon V-0 \to \hat{V}$. It follows from Lemmas A6, A9, A10 and Remark A5 that $\bar{L}$ and $\bar{V} \subset \bar{L}$ are projective, rationally smooth varieties and that

$$|\hat{V}|_r = \frac{N_{y,w}(p^r)\cdot p^{-rl(y)} - 1}{p^r - 1},$$

for all $r \geqq 1$. It is easy to see that the function

$$\alpha(q) \overset{\text{def}}{=} \frac{N_{y,w}\cdot q_y^{-1} - 1}{q-1}$$

is a polynomial in q. By the Lefschetz fixed point formula, we have

$$\alpha(p^r) = \sum_{i=0}^{2d} (-1)^i \operatorname{Tr}(F^r, H^i(\hat{V}))$$

where $d = l(w) - l(y) - 1 = \dim(\hat{V})$ and F is the Frobenius map relative to the F_p-structure. $\hat{V}$ is rationally smooth, projective, hence it satisfies the Weil conjecture. (P. Deligne, La conjecture de Weil, II.) It follows that $H^{2i+1}(\hat{V}) = 0$ for all i and that all eigenvalues of F on $H^{2i}(\hat{V})$ are equal to p^i.

Using the Leray spectral sequence for $\pi\colon V-0 \to \hat{V}$, we get an exact sequence

$$\dots \to H^{i+1}(V-0) \to H^i(\hat{V})(-1) \xrightarrow{\wedge\omega} H^{i+2}(\hat{V}) \xrightarrow{\pi^*} H^{i+2}(V-0) \to \dots \tag{1}$$

where $\omega \in H^2(\hat{V})(1)$ is the restriction oof the corresponding class in $H^2(\hat{L})(1)$.

Let

$$\Pi_{2i} = \operatorname{coker}(H^{2i-2}(\hat{V})(-1) \xrightarrow{\wedge\omega} H^{2i}(\hat{V})),$$

$$\Pi'_{2i} = \ker(H^{2i}(\hat{V})(-1) \xrightarrow{\wedge\omega} H^{2i+2}(\hat{V})).$$

By Poincaré duality on $\hat{V}$, we have

$$\dim \Pi_{2i} = \dim \Pi'_{2d-2i}. \tag{2}$$

Since $H^{2i+1}(\hat{V}) = 0$, it follows from (1) and Lemma A7 that $H'_{2i} \cong H^{2i+1}(V-0) \cong H^{2i+2}_{\langle 0\rangle}(V)$ and that, for $i \neq 0$, $\Pi_{2i} \cong H^{2i}(V-0) \cong H^{2i+1}_{\langle 0\rangle}(V)$, for $i \neq 0$. Hence,

$$\dim \Pi'_{2i} = \dim H^{2i+2}_{\langle 0\rangle}(V) = \dim H^{2l(y)+2i+2}_{\langle b\rangle}(U) = \dim H^{2l(y)+2i+2}_{\langle b\rangle}(\overline{\mathscr{B}_w}) \tag{3}$$

$$\dim \Pi_{2i} = \dim H^{2i+1}_{\langle 0\rangle}(V) = \dim H^{2l(y)+2i+1}_{\langle b\rangle}(U) = \dim H^{2l(y)+2i+1}_{\langle b\rangle}(\overline{\mathscr{B}_w}), \quad \text{for } i \neq 0. \tag{4}$$

We have:

$$\begin{aligned} N_{y,w}(p^r)\cdot p^{-rl(y)} - 1 &= |\hat{V}|_r(p^r-1) \\ &= \sum_i \operatorname{Tr}(F^r, H^{2i}(\hat{V}))(p^r-1) \\ &= \sum_i (\operatorname{Tr}(F^r, H^{2i}(\hat{V})(-1)) - \operatorname{Tr}(F^r, H^{2i+2}(\hat{V}))) \\ &= \sum_i (\operatorname{Tr}(F^r, \Pi'_{2i}) - \operatorname{Tr}(F^r, \Pi_{2i})) \\ &= \sum_i p^{r(i+1)} \dim \Pi'_{2i} - \sum_i p^{ri} \dim \Pi_{2i} \\ &= \sum_i p^{r(d-i+1)} \dim \Pi_{2i} - \sum_i p^{ri} \dim \Pi_{2i}, \quad \text{(by (2)).} \end{aligned}$$

Since this is true for all r, we have an identity of polynomials in q:

$$N_{y,w}\cdot q_y^{-1} - 1 = \sum_i q_w q_y^{-1} q^{-i} \dim \Pi_{2i} - \sum_i q^i \dim \Pi_{2i}.$$

On the other hand, in the proof of Lemma 2.6(ii), we have seen that

$$N_{y,w}\cdot q_y^{-1} - 1 = \overline{P_{y,w}}\cdot q_w q_y^{-1} - P_{y,w}.$$

It follows that

$$q_w^{-1/2} q_y^{1/2}(P_{y,w} - \sum_i q^i \cdot \dim \Pi_{2i}) = \overline{q_w^{-1/2} q_y^{1/2}(P_{y,w} - \sum_i q^i \cdot \dim \Pi_{2i})}. \tag{5}$$

By the Lefschetz theorem [5, Éxp. 13] for V, we have $\dim \Pi_{2i} = 0$ if $i > d/2 = 1/2(l(w) - l(y) - 1)$. It follows that the left hand side of (5) is a polynomial in $q^{-1/2}$ without constant term, hence it cannot be fixed by the involution $a \to \bar{a}$ unless it is zero. Thus, we have

$$P_{y,w} = \sum_{i\geqq 0} q^i \dim \Pi_{2i} = 1 + \sum_{i \geqq 1} q^i \dim H^{2l(y)+2i+1}_{\langle b\rangle}(\overline{\mathscr{B}_w}) \tag{6}$$

and

$$P_{y,w} = \sum_{i\geqq 0} q^i \dim \Pi'_{2d-2i} = \sum_{i\geqq 0} q^i \dim H^{2l(w)-2i}_{\langle b\rangle}(\overline{\mathscr{B}_w}). \tag{7}$$

Moreover, it follows from Lemma A7 that $H^j_{\langle b\rangle}(\overline{\mathscr{B}_w}) = 0$ for $j \leqq 2l(y)+1$. Note also that $\dim H^i_{\langle x\rangle}(\overline{\mathscr{B}_w})$ is constant when x runs through $\mathscr{B}_y$. Using now (6) and (7) it follows directly that $P_{y,w} = 1$ if and only if $\mathscr{B}_y \cap \mathscr{S}(\overline{\mathscr{B}_w}) = 0$, and Theorem A2 is proved.

Remark A 11. In the process of proving Theorem 2, we have also obtained the explicit formulae (6), (7) for $P_{y,w}$, valid for any $y<w$ in W such that $P_{y',w}=1$ for all $y<y'\leqq w$. From this, we see that for such pairs $y<w$, we have $y\prec w$ if and only if (with the notations of the previous proof) we have $d\equiv 0 \pmod 2$ and $\dim \Pi_d \neq 0$ (i.e. if $\hat{V}$ has non trivial "primitive cohomology" in the middle dimension).

Corollary A 12. *For any $w\in W$, $\mathscr{S}(\overline{\mathscr{B}_w})$ has codimension $\geqq 3$ in $\overline{\mathscr{B}_w}$.*

(See Lemma 2.6(iii), (iv).)

This is in contrast with the behaviour of the singular set of $\overline{\mathscr{B}_w}$. For example, if $G=Sp_4$ and w, w' are the two elements of length 3 in W, then one of the Schubert cells $\mathscr{B}_w, \mathscr{B}_{w'}$ is non-singular, and the other one has a singular set of codimension 2.

References

1. Benson, C.T., Curtis, C.W.: On the degrees and rationality of certain characters of finite Chevalley groups, Trans. Amer. Math. Soc. **165**, 251–273 (1972) and **202**, 405–406 (1975)
2. Bourbaki, N.: Groupes et algèbres de Lie, Chapitres 4, 5, 6, Paris: Hermann 1968
3. Curtis, C.W., Iwahori, N., Kilmoyer, R.: Hecke algebras and characters of parabolic type of finite groups with (B,N) pair, I.H.E.S. Publ. Math. **40**, 81–116 (1972)
4. Delorme, P.: Extensions dans la categorie $\mathcal{O}$ de Bernstein-Gelfand-Gelfand. Applications, preprint, Palaiseau, 1978
5. Grothendieck, A., SGA2, Amsterdam: North Holland 1968
6. Jantzen, C.J.: Moduln mit einem höchsten Gewicht. Habilitationsschrift, Bonn, 1977
7. Joseph, A.: A characteristic variety for the primitive spectrum of a semisimple Lie algebra, preprint, Bonn, 1976
8. Joseph, A.: W-module structure in the primitive spectrum of the enveloping algebra of a semisimple Lie algebra, preprint, Tel-Aviv, 1978
9. Knuth, D.E.: The art of computer programming, Addison-Wesley, 1975
10. Lusztig, G.: A class of irreducible representations of a Weyl group, in press (1979)
11. Verma, D.N.: Möbius inversion for the Bruhat ordering on a Weyl group, Ann. Sci. E.N.S., 4[e] serie, t. **4**, 393–398 (1971)
12. Vogan, D.: A generalized τ-invariant for the primitive spectrum of a semisimple Lie algebra, preprint, Princeton, 1978
13. Vogan, D.: Ordering in the primitive spectrum of a semisimple Lie algebra, preprint, Princeton, 1978
14. Vogan, D.: Irreducible characters of semisimple Lie groups II. The Kazhdan-Lusztig conjectures, manuscript, 1979

Received March 11, 1979

Inventiones math. 53, 185–191 (1979)

Inventiones mathematicae

The Finiteness Obstructions for Nilpotent Spaces lie in $D(\mathbb{Z}\pi)$

G. Mislin[1] and K. Varadarajan[2]*

[1] Mathematisches Institut der ETH, Zürich, Switzerland
[2] The University of Calgary, Calgary, Canada

Introduction

Let X be a finitely dominated nilpotent space and $\pi=\pi_1(X)$. In [9] it is shown that the unreduced Wall obstruction $w(X)$ is zero whenever π is infinite. In [8] results are obtained concerning the restrictions that $w(X)$ should satisfy in case π is a finite p-group, with p a prime. Let π be finite, $\overline{\mathbb{Z}\pi}$ a maximal $\mathbb{Z}$ order in $\mathbb{Q}\pi$ containing $\mathbb{Z}\pi$ and $D(\mathbb{Z}\pi)=\operatorname{Ker}(j_*\colon K_0(\mathbb{Z}\pi)\to K_0(\overline{\mathbb{Z}\pi}))$ where $j\colon \mathbb{Z}\pi\to\overline{\mathbb{Z}\pi}$ denotes the inclusion. One of the main results in [8] is that if π is a cyclic group of order p^k then $w(X)\in D(\mathbb{Z}\pi)$. In [13] this result was extended to the case when π is any finite abelian group. The object of the present paper is to establish this result for all finite groups π.

For an arbitrary group π, using nilpotent (resp. special FP) complexes we define two subgroups $N(\mathbb{Z}\pi)$ (resp. $S(\mathbb{Z}\pi)$) of $K_0(\mathbb{Z}\pi)$. Actually the group $N(\mathbb{Z}\pi)$ has already been introduced in [10] by one of the authors in connection with studying those elements of $K_0(\mathbb{Z}\pi)$ which are realizable as Wall obstructions of finitely dominated nilpotent spaces. It will turn out that $N(\mathbb{Z}\pi)\subset S(\mathbb{Z}\pi)$ whenever π is finitely generated nilpotent, and $S(\mathbb{Z}\pi)\subset N(\mathbb{Z}\pi)$ whenever π is finite. In terms of these groups our main result is that $S(\mathbb{Z}\pi)=N(\mathbb{Z}\pi)\subset D(\mathbb{Z}\pi)$ whenever π is finite and nilpotent. Finally we give an example to show that in general $N(\mathbb{Z}\pi)\neq D(\mathbb{Z}\pi)$ even when π is finite cyclic.

All the spaces we consider are pointed connected CW-complexes. $\tilde{X}$ will denote the universal covering of X. We will refer to chain complexes merely as complexes and throughout we consider only positive chain complexes.

1. Special Complexes, Nilpotent Complexes and Perfect Modules

Let π denote an arbitrary group and $I\subset\mathbb{Z}\pi$ the augmentation ideal of $\mathbb{Z}\pi$. Recall that a π-module M is nilpotent if $I^k M=0$ for some $k>0$, and M is perfect if $IM=M$. A similar definition for right π-modules.

* Research done while the second author was partially supported by NRC grant A 8225

0020-9910/79/0053/0185/$01.40

Definition 1.1. A π-complex C_* will be said to be nilpotent, if $H_i(C_*)$ are nilpotent π-modules for all i.

For instance if X is a homologically nilpotent space, then by definition, the singular chain complex $S_*\tilde{X}$ is a nilpotent $\pi_1(X)$-complex.

Theorem 1.2. *Let π be a nilpotent group and $f\colon \pi \to \bar{\pi}$ a surjective map of groups. Suppose C_* is a projective, nilpotent π-complex. Then $f_* C_* = \mathbb{Z}\bar{\pi} \otimes_{\pi} C_*$ is a projective, nilpotent $\bar{\pi}$-complex.*

Proof. It is clear that $f_* C_*$ is a projective $\bar{\pi}$-complex. To see that $H_i(f_* C_*) = H_i(C_*, \mathbb{Z}\bar{\pi})$ is $\bar{\pi}$-nilpotent we consider the spectral sequence $\operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, H_j(C_*)) \Rightarrow H_{i+j}(C_*, \mathbb{Z}\bar{\pi})$ [Theorem 5.5.1, Chap. I in [5]]. It sufficies to show that $\operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, H_j(C_*))$ is a nilpotent $\bar{\pi}$-module for all i and j. First we consider $\operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, \mathbb{Z})$ with trivial π-action on $\mathbb{Z}$. Let $K = \operatorname{Ker}(f\colon \pi \to \bar{\pi})$. Then we have the short exact sequence $K \to \pi \xrightarrow{f} \bar{\pi}$ of nilpotent groups and $\operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, \mathbb{Z}) = H_i(\pi, \mathbb{Z}\bar{\pi}) \cong H_i(K, \mathbb{Z})$ as $\bar{\pi}$-modules. It is known that $\bar{\pi}$ acts nilpotently on $H_i(K, \mathbb{Z})$. Let $(I\bar{\pi})^{k_i} \operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, \mathbb{Z}) = 0$. If A is any finitely generated abelian group which trivial π-action, we can get an exact sequence $0 \to \mathbb{Z}^s \to \mathbb{Z}^r \to A \to 0$ for suitable integers $r, s \geqq 0$. From the exactness of

$$\cdots \to \operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, \mathbb{Z}^s) \to \operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, \mathbb{Z}^r) \to \operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, A) \to \operatorname{Tor}_{i-1}^{\pi}(\mathbb{Z}\bar{\pi}, \mathbb{Z}^s) \to \cdots$$

we see that $(I\bar{\pi})^{k_i + k_{i-1}} \operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, A) = 0$. Since $\operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, -)$ commutes with direct limits we see that $(I\bar{\pi})^l \operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, B) = 0$ for any abelian group B where $l = k_i + k_{i+1}$. If M is any nilpotent π-module, an easy induction on the nilpotency class of M shows that $\operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, M)$ is $\bar{\pi}$-nilpotent. Since each $H_j(C_*)$ is a nilpotent π-module we see that $\operatorname{Tor}_i^{\pi}(\mathbb{Z}\bar{\pi}, H_j(C_*))$ is $\bar{\pi}$-nilpotent.

Theorem 1.3. *Let π denote a finitely generated nilpotent group, C_* a projective, nilpotent π-complex and M a finitely generated perfect right π-module. Then $H_i(C_*, M) = H_i(M \otimes_{\pi} C_*) = 0$ for all i.*

Proof. From Theorem 3 of [1] we see that $\operatorname{Tor}_i^{\pi}(M, \mathbb{Z}) = H_i(\pi, M) \cong H_i(\pi, \hat{M}) = 0$, since the I-adic completion $\hat{M}$ of a perfect module is zero. Using the same kind of arguments as in the proof of Theorem 1.2, we infer that $\operatorname{Tor}_i^{\pi}(M, C) = 0$ for any nilpotent π-module C. Theorem 1.3 is now immediate from the spectral sequence $\operatorname{Tor}_i^{\pi}(M, H_j(C_*)) \Rightarrow H_{i+j}(C_*, M)$.

Remark. There is an obvious "local version" of Theorem 1.3, assuming that C_* is a $\mathbb{Z}_{(p)}\pi$-projective, nilpotent π-complex and M a finitely generated right $\mathbb{Z}_{(p)}\pi$-module which is π-perfect. Here $\mathbb{Z}_{(p)}$ denotes the localization of $\mathbb{Z}$ at a prime p. One uses the fact that $I\mathbb{Z}_{(p)}\pi$ has the Artin-Rees property whenever π is finitely generated nilpotent.

Recall that a π-complex C_* is called special [13] if for all $H \subset G \subset \pi$ with $(G:H) < \infty$, the canonical map $H_*(\mathbb{Z}_{(p)} \otimes_H C_*) \to H_*(\mathbb{Z}_{(p)} \otimes_G C_*)$ is an isomorphism for all primes p which do not divide $(G:H)$.

Corollary 1.4. *Let π be a finitely generated nilpotent group and C_* a projective, nilpotent π-complex. Then C_* is special.*

Proof. Let $H \subset G \subset \pi$, $(G:H) < \infty$ and p a prime not dividing $(G:H)$. In order to show that $H_*(\mathbb{Z}_{(p)} \underset{H}{\otimes} C_*) \to H_*(\mathbb{Z}_{(p)} \underset{G}{\otimes} C_*)$ is a isomorphism, we have only to prove that $H_*(K \underset{\pi}{\otimes} C_*) = 0$ where $K =$ (kernel of the canonical quotient map $\mathbb{Z}_{(p)} \underset{H}{\otimes} \mathbb{Z}\pi \to \mathbb{Z}_{(p)} \underset{G}{\otimes} \mathbb{Z}\pi$). K is clearly a finitely generated $\mathbb{Z}_{(p)}\pi$-module. Moreover $K \underset{\pi}{\otimes} C_* \cong (K \underset{\pi}{\otimes} \mathbb{Z}_{(p)}\pi) \underset{\pi}{\otimes} C_* \cong K \underset{\pi}{\otimes} (\mathbb{Z}_{(p)}\pi \underset{\pi}{\otimes} C_*)$. Clearly $\mathbb{Z}_{(p)}\pi \underset{\pi}{\otimes} C_*$ is a $\mathbb{Z}_{(p)}\pi$-projective complex. By the local version of Theorem 1.3, it suffices to show that K is perfect. The exact sequence $0 \to K \to \mathbb{Z}_{(p)} \underset{H}{\otimes} \mathbb{Z}\pi \to \mathbb{Z}_{(p)} \underset{G}{\otimes} \mathbb{Z}\pi \to 0$ of right π-modules yields the exact sequence

$$\cdots \to H_1(H, \mathbb{Z})_{(p)} \to H_1(G, \mathbb{Z})_{(p)} \to H_0(\pi, K) \to 0$$

(where we identify $H_1(\pi, \mathbb{Z}_{(p)} \underset{H}{\otimes} \mathbb{Z}\pi)$ with $H_1(H, \mathbb{Z})_{(p)}$ etc.). Since p is prime to $(G:H)$ the inclusion $H \hookrightarrow G$ induces isomorphisms $H_{(p)} \cong G_{(p)}$. From Theorem 1.12, Chap. I of [7] we see that $\tilde{H}_*(H)_{(p)} \cong \tilde{H}_*(G)_{(p)}$. In particular $H_1(H, \mathbb{Z})_{(p)} \cong H_1(G, \mathbb{Z})_{(p)}$ and hence $H_0(\pi, K) = 0$. Thus $K = KI$ is perfect.

Recall that in case M is a left π-module, the homology groups of π with coefficients in M are defined to be $\mathrm{Tor}_i^\pi(Z, M)$ where Z is the trivial right $\mathbb{Z}\pi$-module $\mathbb{Z}$. It has been proved in [13] that if π is a finite group and C_* is a special FP complex, then every p-Sylow subgroup $\pi_p \subset \pi$ acts nilpotently on $H_*(C_*)$. Let q be any prime $\neq p$ and M any nilpotent $\mathbb{Z}_{(q)}\pi_p$-module. Let I be the augmentation ideal in $\mathbb{Z}\pi_p$. Then M and M/IM are both q-local from which it is immediate that $H_i(\pi_p, M)$ and $H_i(\pi_p, M/IM)$ are all q-local. From Theorem 4.12, Chap. I of [7] it follows that $H_i(\pi_p, M) \cong H_i((\pi_p)_q, M)$ and $H_i(\pi_p, M/IM) \cong H_i((\pi_p)_q, M/IM)$ for all $i \geqq 0$. Since $(\pi_p)_q = \{1\}$ we see that $H_i(\pi_p, M) = 0 = H_i(\pi_p, M/IM)$ for all $i \geqq 1$. In the exact homology sequence

$$\cdots \to H_1(\pi_p, M/IM) \to H_0(\pi_p, IM) \to H_0(\pi_p, M) \to H_0(\pi_p, M/IM) \to 0$$

corresponding to the exact sequence $0 \to IM \to M \to M/IM \to 0$ we have $H_0(\pi_p, M) \cong H_0(\pi_p, M/IM)$ both being isomorphic to M/IM. In fact $I(M/IM) = 0$ and hence $H_0(\pi_p, M/IM) = M/IM$. Hence we get $H_1(\pi_p, M/IM) \cong H_0(\pi_p, IM)$. But $H_1(\pi_p, M/IM) = 0$ and $H_0(\pi_p, IM) \cong IM/I^2M$. It follows that $IM = I^2M$ and hence $IM = I^kM$ for all $k \geqq 1$. Since M is nilpotent we see that $IM = 0$. Hence M is a trivial π_p-module.

Let C_* be special FP complex over a finite group π. Then from the above comments it follows that for any prime p, any Sylow p-subgroup π_p acts nilpotently on $H_*(C_*)$ and hence acts trivially on $H_*(C_*)_{(q)}$ for any prime $q \neq p$. Thus for any prime q, every Sylow p-subgroup π_p of π acts *trivially* on $H_*(C_*)_{(q)}$ whenever $p \neq q$, and π_q itself acts nilpotently on $H_*(C_*)_{(q)}$. Hence the action of π on $H_*(C_*)_{(q)}$ is nilpotent. Now $H_*(C_*) \subset \prod_q H_*(C_*)_{(q)}$, the product taken over all primes and moreover, except for those primes q which divide the order of π the action of π on $H_*(C_*)_{(q)}$ is trivial. For primes q which divide the order of π, the action of π on $H_*(C_*)_{(q)}$ is nilpotent. Hence $\prod_q H_*(C_*)_{(q)}$ is a nilpotent π-module and so is $H_*(C_*)$. Hence we have

Proposition 1.5. *If C_* is a special FP-complex over a finite group π then C_* is a nilpotent complex.*

Corollary 1.6. *If π is a finite nilpotent group, then the following are equivalent.*

(i) *C_* is a special FP-complex*

(ii) *C_* is a nilpotent FP-complex.*

This follows from Corollary 1.4 *and Proposition* 1.5.

2. The Subgroups $S(\mathbb{Z}\pi)$ and $N(\mathbb{Z}\pi)$ of $K_0(\mathbb{Z}\pi)$

Definition 2.1. For a nontrivial group π define $S(\mathbb{Z}\pi)$ (resp. $N(\mathbb{Z}\pi)$) to be the subgroup of $K_0(\mathbb{Z}\pi)$ generated by the elements x which may be represented in the form $x=\Sigma(-1)^i[C_i]\in K_0(\mathbb{Z}\pi)$, where $C_*=\{C_i,\partial_i\}$ is a special (resp. nilpotent) complex of type FP. If $\pi=\{1\}$ define $S(\mathbb{Z}\pi)=N(\mathbb{Z}\pi)=0$.

Observe that $S(\mathbb{Z}-)$ and $N(\mathbb{Z}-)$ are functors on certain categories of groups and surjective maps. For $N(\mathbb{Z}-)$ this is a consequence of Theorem 1.2 and for $S(\mathbb{Z}-)$ it is a consequence of Proposition 3.6 of [13].

If π is finitely generated nilpotent, then $N(\mathbb{Z}\pi)\subset S(\mathbb{Z}\pi)$ by Corollary 1.4. If π is arbitrary but finite, then $S(\mathbb{Z}\pi)\subset N(\mathbb{Z}\pi)$ by Proposition 1.5. Hence we have

Theorem 2.2. *Let π be a finite nilpotent group. Then $S(\mathbb{Z}\pi)=N(\mathbb{Z}\pi)$.*

For more results concerning the structure of $N(\mathbb{Z}\pi)$ the reader may refer to [10].

Let π be a finite group. Denote by $D(\mathbb{Z}\pi)\subset K_0(\mathbb{Z}\pi)$ the kernel of j_*: $K_0(\mathbb{Z}\pi)\to K_0(\overline{\mathbb{Z}\pi})$ where $j\colon \mathbb{Z}\pi\to\overline{\mathbb{Z}\pi}$ denotes the inclusion of $\mathbb{Z}\pi$ in a maximal $\mathbb{Z}$-order $\overline{\mathbb{Z}\pi}$ in $\mathbb{Q}\pi$ containing $\mathbb{Z}\pi$. The main result of this paper is the following.

Theorem 2.3. *Let π be a finite nilpotent group. Then $S(\mathbb{Z}\pi)=N(\mathbb{Z}\pi)\subset D(\mathbb{Z}\pi)$.*

Let $\overline{\mathbb{Z}\pi}$ be a maximal $\mathbb{Z}$-order in $\mathbb{Q}\pi$ with $\mathbb{Z}\pi\subset\overline{\mathbb{Z}\pi}$. If $\mathbb{Q}\pi=L_0\times L_1\times\ldots\times L_r$ with each L_i a simple algebra over $\mathbb{Q}$, then $\overline{\mathbb{Z}\pi}=A_0\times A_1\times\ldots\times A_r$ with each A_i a maximal $\mathbb{Z}$-order in L_i. If $N=\sum_{x\in\pi}x$, then $\mathbb{Q}\cong\mathbb{Q}N$ is a simple component of $\mathbb{Q}\pi$. We can assume $L_0=\mathbb{Q}N$ and thus $A_0\cong\mathbb{Z}$. Let $f_i\colon \mathbb{Z}\pi\to A_i$ be the composition of j: $\mathbb{Z}\pi\hookrightarrow\overline{\mathbb{Z}\pi}$ with the projection $A_0\times\ldots\times A_r\to A_i$. To prove that $S(\mathbb{Z}\pi)\subset D(\mathbb{Z}\pi)$ we have only to show that $(*)$: $f_{i_*}S(\mathbb{Z}\pi)=0\subset K_0(A_i)$, $1\leqq i\leqq r$. Since $S(\mathbb{Z}-)$ is a functor on the category of groups and epimorphisms, it suffices to prove $(*)$ in the case when f_i is faithful, namely in the case when $f_i|\pi$ is $(1-1)$. Thus we will actually assume that f_i is faithful. We distinguish three cases:

(1) π is a p-group, with p an odd prime

(2) π is a 2-group, and

(3) π is a group of mixed order.

Case (1). In this case $\mathbb{Q}\otimes_{\mathbb{Z}}A_i$ is a full matrix algebra say $M_n(k)$ with $k=\mathbb{Q}(\omega)$ for some p^m-th root of unity ω, and we can assume $A_i=M_n(\mathbb{Z}[\omega])$. We write A for A_i and f for f_i: $\mathbb{Z}\pi\to A_i$. From Morita's theorems we see that the map μ:

$K_0 M_n(\mathbb{Z}[\omega]) \to K_0(\mathbb{Z}[\omega])$ given by $\mu([P]) = [\mathbb{Z}[\omega]^n \underset{M_n(\mathbb{Z}[\omega])}{\otimes} P]$ is an isomorphism. Let $x = \Sigma(-1)^i [P_i]$ with P_* a special FP-complex over $\mathbb{Z}\pi$. Since $|\pi| \overline{\mathbb{Z}\pi} \subset \mathbb{Z}\pi$, the arguments used in the proof of Lemma 3.4 of [8] show that $H_*(A \underset{\pi}{\otimes} P_*)$ is a finite p-torsion abelian group. Since any Morita equivalence preserves exactness, it follows that $H_*(\mathbb{Z}[\omega]^n \underset{A}{\otimes} A \underset{\pi}{\otimes} P_*) = \mathbb{Z}[\omega]^n \underset{A}{\otimes} H_*(A \otimes P_*)$ is itself a p-torsion, finite abelian group. We identify $K_0(\mathbb{Z}[\omega])$ with $G_0(\mathbb{Z}[\omega])$ using the Cartan map. If we show that any $\mathbb{Z}[\omega]$-module, which is finite and p-torsion as an abelian group, represents 0 in $G_0(\mathbb{Z}[\omega])$ then it will follow that $\mu f_*(x) = 0$. Using the exact sequence $\mathbb{Z}[\omega] \xrightarrow{1-\omega} \mathbb{Z}[\omega] \twoheadrightarrow \mathbb{Z}/p\mathbb{Z}$, it is already shown in [8] that any $\mathbb{Z}[\omega]$ module which is finite and p-torsion represents 0 in $G_0(\mathbb{Z}[\omega])$. Hence $\mu f_*(x) = 0$. Since μ is monic, we get $f_*(x) = 0$.

Case (2). π is a 2-group. During the course of this proof we will be following the notations used in [2]. By a result of *W.* Feit (Theorem 14.3 of [4]) we may assume that the faithful irreducible representation $\mathbb{Q} \otimes f_i\colon \mathbb{Q}\pi \to \mathbb{Q} \otimes A_i = \Sigma_i$ is induced from a representation $\mathbb{Q}\pi_i \to \Sigma_i'$ of a subgroup $\pi_i \subset \pi$ such that $E_i = \pi_i/\mathrm{Ker}(\pi_i \to \Sigma_i')$ is a special elementary group and $\Sigma_i = \mathrm{End}_{\Sigma_i'}(\mathbb{Q}\pi \underset{\pi_i}{\otimes} \Sigma_i')$; (see the proof of Theorem 2.4 of [2]; for the definition of a special elementary group see [2], section 1). Following [2] we can identify Σ_i' with $\Sigma(E_i)$, the unique E_i-faithful simple component of $\mathbb{Q}E_i$. The symbols $K(E_i)$, $R(E_i)$ will have the same meaning as in [2] ($K(E_i)$ is the center of $\Sigma(E_i)$ and $R(E_i)$ the ring of algebraic integers of $K(E_i)$). Let $\Omega(E_i)$ be a maximal $R(E_i)$-order in $\Sigma(E_i)$. Then $\Omega_i = \mathrm{End}_{\Omega(E_i)}(\mathbb{Z}\pi \underset{\pi_i}{\otimes} \Omega(E_i))$ is a maximal $R(E_i)$-order in Σ_i. We may assume $A_i = \Omega_i$. Since $\mathbb{Z}\pi$ is free over $\mathbb{Z}\pi_i$ of rank$d = (\pi:\pi_i)$ we may regard $\mathrm{End}_{\Omega(E_i)}(\mathbb{Z}\pi \underset{\pi_i}{\otimes} \Omega(E_i))$ as $M_d(\Omega(E_i))$. Hence we get an isomorphism μ: $K_0(A_i) \to K_0(\Omega(E_i))$ given by $[P] \mapsto [\Omega(E_i)^d \underset{A_i}{\otimes} P]$. Since E_i is a special elementary 2-group, it will be one of C_{2^l}, D_{2^l}, SC_{2^l}, SD_{2^l} or H_{2^l} in the notation of [2]. We now distinguish between two sub-cases.

Sub-case (a): $E_i \neq H_{2^l}$.

In this case, by Lemma 1.1 of [2], $\Sigma(E_i) \cong M_n(K(E_i))$ for some n. Since $\Omega(E_i)$ is a maximal $R(E_i)$ order in $\Sigma(E_i)$, we may assume $\Omega(E_i) = M_n(R(E_i))$. Let ν: $K_0(\Omega(E_i)) \to K_0(R(E_i))$ be the isomorphism $[X] \mapsto [R(E_i)^n \underset{\Omega(E_i)}{\otimes} X]$ arising from Morita equivalence. Then $K_0(A_i) \xrightarrow{\nu\mu} K_0(R(E_i))$ is an isomorphism. If P_* is a special FP-complex over $\mathbb{Z}\pi$, using $|\pi| \overline{\mathbb{Z}\pi} \subset \mathbb{Z}\pi$, one sees that $H_*(A_i \underset{\pi}{\otimes} P_*)$ is a finite, 2-torsion group. Following the arguments used in the proof of case (1) and noting that $K_0(R(E_i)) \cong G_0(R(E_i))$ since $R(E_i)$ is a ring of algebraic integers, to show that $f_{i_*}(S(\mathbb{Z}\pi)) = 0$ we have only to prove that any $R(E_i)$-module which is finite, 2-torsion represents 0 in $G_0(R(E_i))$. This follows from the fact that $R(E_i) \xrightarrow{\alpha_{E_i}} R(E_i) \twoheadrightarrow \mathbb{Z}/2\mathbb{Z}$ is exact, where α_{E_i} is the element of $R(E_i)$ described in [2, Sect. 1].

Sub-case (b): $E_i = H_{2^l}$ (generalized quaternion group of order 2^{l+1}).

Before proceeding with this case, first observe that Lemma 3.4 of [8] actually yields $w(X)\in D(\mathbb{Z}\pi)$ if X is a finitely dominated nilpotent space with $\pi=\pi_1(X)$ a cyclic p-group. Hence in Lemma 4.1 of [13] we can replace $\tilde{w}(C)$ by $w(C)$. In particular if π is a 2-group and P_* a special FP-complex over π, then $\Sigma(-1)^i\,[P_i]$ is in the 2-torsion of $K_0(Z\pi)$. Thus $S(\mathbb{Z}\pi)\subset$2-torsion of $K_0(Z\pi)$, whenever π is a 2-group. If we show that $\tilde{K}_0 A_i$ has odd order in this case, it will follow that $S(\mathbb{Z}\pi)\subset D(\mathbb{Z}\pi)$. We still have $A_i=\Omega_i=\mathrm{End}_{\Omega(E_i)}(\mathbb{Z}\pi\underset{E_i}{\otimes}\Omega(E_i))$. From Morita, we see that $K_0(A_i)\rightarrowtail K_0\Omega(E_i)$. Here $\Omega(E_i)$ is a maximal $R(E_i)$-order in $\mathbb{Q}\underset{\mathbb{Z}}{\otimes}\Omega(E_i)$ which is itself a simple component of $\mathbb{Q}E_i$. Observe that $E_i=H_{2^l}$ with $l\geqq 2$. It is known in this case [3] that the (locally free) class group $Cl(\mathbb{Z}E_i)$ is isomorphic to $\bigoplus_{j=1}^{l-1} Cl(\Gamma_j)$ where $\Gamma_j=$the ring of real integers in $\mathbb{Q}(\sqrt[2^j]{1})$. A result of Weber [6] asserts that $Cl(\Gamma_j)$ is of odd order. Since $\tilde{K}_0(\Omega(E_i))$ is a subgroup of $Cl(\overline{\mathbb{Z}E}_i)$ we see that $\tilde{K}_0(A_i)$ is also of odd order. This completes the proof of subcase (b).

Case (3). π is of mixed order. As always, $\mathbb{Q}\underset{\mathbb{Z}}{\otimes}A_i$ is a central simple algebra, with centre k say. The map $f_i\colon \mathbb{Z}\pi\to A_i\subset\mathbb{Q}\underset{\mathbb{Z}}{\otimes}A_i$ maps the centre of π into k. Since π is nilpotent, it is isomorphic to the direct product of its Sylow subgroups. Each non-trivial Sylow subgroup has a non-trivial centre. Hence we can find an element $x\in$(centre of π) with x having a mixed order, say m. Then $f_i(x)\in k$. Let Φ_m denote the cyclotomic polynomial of order m. Since f_i is faithful by assumption, we see that $f_i(x)\in k$ is a primitive m-th root of unity and hence if $\sigma=\exp\left(\frac{2\pi\sqrt{-1}}{m}\right)$ we have an obvious commutative diagram

$$\begin{tikzcd} Z\pi \arrow[r, "f_i"] & A_i \\ Z[x] \arrow[u] \arrow[r, two heads] & Z[\sigma] \arrow[u] \end{tikzcd}$$

where σ gets mapped into $f_i(x)\in A_i$. Since σ is of mixed order, $1-\sigma$ is a unit in $\mathbb{Z}[\sigma]$, and it follows that $f_i(1-x)$ is a unit in A_i. Hence A_i is a (right) π-perfect module. From Corollary 1.6 and Theorem 1.3 we see that for any special FP-complex P_* over π, we have $H_*(A_i\underset{\pi}{\otimes}P_*)=0$. Since $K_0(A_i)\cong G_0(A_i)$, we see immediately that $f_{i_*}(\Sigma(-1)^i\,[P_i])=0$. Thus $f_{i_*}(S(\mathbb{Z}\pi))=0$. This completes the proof of Theorem 2.3.

Remark. The inclusion $S(\mathbb{Z}\pi)\subset D(\mathbb{Z}\pi)$ was proved in [13] for all finite abelian groups π.

Corollary 2.4. *Let X be a finitely dominated nilpotent space with finite nontrivial fundamental group $\pi_1(X)=\pi$. Then*

(i) $w(X)\in D(\mathbb{Z}\pi)$.

(ii) *If π is a p-group of order $\geqq p^2$, then $p^{-1}|\pi|\,w(X)=0$ if p is odd, and $2^{-2}|\pi|\,w(X)=0$ if $p=2$.*

Proof. (i) is immediate from Theorem 2.3.

(ii) follows from the corresponding result on $D(\mathbb{Z}\pi)$ [12].

Similarly, one can obtain other results on $w(X)$ whenever one knows something about $D(\mathbb{Z}\pi)$. For instance $D(\mathbb{Z}\pi)=0$ whenever π is a dihedral 2-group, and $D(\mathbb{Z}\pi)$ is of order 2 whenever π is a (generalized) quaternion 2-group [3].

The following example shows that in general $N(\mathbb{Z}\pi)\neq D(\mathbb{Z}\pi)$ even when π is finite cyclic. Let π be a cyclic group of order 22. Then from (ii) Theorem 8.15 in [11] we see that $N(2,11)=3$ divides the order of $D(\mathbb{Z}\pi)$. Also from [11] we have $|D(\mathbb{Z}\pi)|>11^{(11\lambda/\log 11)-1}$ for some positive constant λ. Hence $D(\mathbb{Z}\pi)\neq 0$. Thus there is an element of order 3 in $D(\mathbb{Z}\pi)$. From Corollary 3.2 of [10] one sees that $N(\mathbb{Z}\pi)=0$.

References

1. Brown, K.S., Dror, E.: The Artin-Rees property and homology: Israel Journal of Mathematics, **22**, 93–109 (1975)
2. Endo, S., Miyata, T.: On the projective class group of finite groups; Osaka J. of Math., **13**, 109–122 (1976)
3. Fröhlich, A., Keating, M.E., Wilson, S.M.J.: The class groups of quaternion and dihedral 2-groups; Mathematika, **21**, 64–71 (1974)
4. Feit, W.: Characters of finite groups, New York: W.A. Benjamin Inc. 1967
5. Godement, R.: Topologie algébrique et théorie des faisceaux, Hermann, Paris 1958
6. Hasse, H.: Über die Klassenzahl Abelscher Zahlkörper; Berlin: Akademie-Verlag 1952
7. Hilton, P.J., Mislin, G., Roitberg, J.: Localization of nilpotent groups and spaces; Mathematics Studies 15, Amsterdam: North-Holland Publishing Company, 1975
8. Mislin, G.: Wall's obstruction for nilpotent spaces; Topology **14**, 311–317 (1975)
9. Mislin, G.: Finitely dominated nilpotent spaces; Annals of Math. **103**, 547–556 (1976)
10. Mislin, G.: The geometric realization of Wall obstructions by nilpotent and simple spaces; preprint
11. Reiner, I., Ullom, S.: A Mayer-Vietoris sequence for class groups; J. Alg. **31**, 305–342 (1974)
12. Ullom, S.: The exponent of class groups; J. Alg. **29**, 124–132 (1974)
13. Varadarajan, K.: Finiteness obstruction for nilpotent spaces; J. Pure and Appl. Algebra, **12**, 137–146 (1978)

Received May 4, 1979

Inventiones math. 53, 193–225 (1979)

Inventiones mathematicae

Plane Partitions (III): The Weak Macdonald Conjecture

George E. Andrews*

The Pennsylvania State University, University Park, Pennsylvania 16802, U.S.A.

Dedicated to the memory of Alfred Young and F.J.W. Whipple

1. Introduction

A plane partition π of n is an array of positive integers a_{ij} such that $\sum a_{ij}=n$ (more briefly $|\pi|=n$) where $a_{ij}\geqq\max(a_{i+1,j}, a_{i,j+1})$. Generally we represent a plane partition by the array

$$\begin{array}{llllll} a_{11} & a_{12} & a_{13} & \cdots & & a_{1r} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2s} & \\ \vdots & & & & & \\ a_{j1} & \cdots & \cdots & \cdots & a_{jw}. & \end{array}$$

Thus the plane partitions of 3 are

$$\begin{array}{llllll} 3, & 21, & 2, & 111, & 11, & 1 \\ & & 1 & & 1 & 1 \\ & & & & & 1. \end{array}$$

There are some well-known theorems and open questions related to plane partitions in which the number of rows and columns and the size of the parts is restricted. I.G. Macdonald [10] has devised a notation that allows a uniform consideration of these questions. First one considers the "Ferrers graph" $D(\pi)$ of a plane partition π; this is the set of integer points (i,j,k) in the first octant that satisfy $1\leqq k\leqq a_{ij}$. Next define the height of $p=(i,j,k)$ to be $ht(p)=i+j+k-2$. Defining

$$\mathscr{B}_{l,m,n}=[1,l]\times[1,m]\times[1,n]\subset\mathbb{Z}^3, \tag{1.1}$$

* Partially supported by National Science Foundation Grant MCS 75-19162

0020-9910/79/0053/0193/$06.60

Macdonald notes [10] that MacMahon's [12; p. 293] main result on plane partitions asserts that

$$\sum_{D(\pi)\subseteq \mathscr{B}_{l,m,n}} q^{|\pi|}=\prod_{p\in \mathscr{B}_{l,m,n}} \frac{1-q^{1+ht(p)}}{1-q^{ht(p)}}. \tag{1.2}$$

Furthermore, MacMahon [11] made a conjecture about symmetric plane partitions. Let G_2 be the two element subgroup of the symmetric group on three letters that contains the identity and the transposition $(i,j,k)\to(j,i,k)$.

Macdonald [10] points out that MacMahon's conjecture reduces to

$$\sum_{\substack{D(\pi)\subseteq \mathscr{B}_{l,l,n}\\ D(\pi)\text{ is } G_2 \text{ invariant}}} q^{|\pi|}=\prod_{\xi\in \mathscr{B}_{l,l,n}/G_2} \frac{(1-q^{|\xi|+ht(\xi)})}{(1-q^{ht(\xi)})}, \tag{1.3}$$

where $|\xi|=\operatorname{Card}(\xi)$, $ht(\xi)=\sum_{p\in\xi} ht(p)$. Macdonald (unpublished) and I [1], [4] have independently proved this conjecture.

Next Macdonald [10] considers G_3, the three element group of cyclic permutations of (i,j,k), and he conjectures

$$\sum_{\substack{D(\pi)\subseteq \mathscr{B}_{m,m,m}\\ D(\pi)\text{ is } G_3 \text{ invariant}}} q^{|\pi|}=\prod_{\xi\in \mathscr{B}_{m,m,m}/G_3} \frac{(1-q^{|\xi|+ht(\xi)})}{(1-q^{ht(\xi)})}. \tag{1.4}$$

While the formulae (1.2), (1.3) and (1.4) beautifully illustrate the parallel nature of these assertions, one becomes aware very quickly that an immense number of cancellations occur in each of these products. Hence we choose to reformulate (1.4) less elegantly but more tractably as follows:

Macdonald's Conjecture. Let $M(m,n)$ denote the number of plane partitions π such that $|\pi|=n$, $D(\pi)$ is invariant under G_3 and $D(\pi)\subseteq\mathscr{B}_{m,m,m}$. Then

$$\sum_{n\geqq 0} M(m,n)\,q^n=\prod_{i=1}^{n}\left\{\frac{(1-q^{3i-1})}{(1-q^{3i-2})}\prod_{j=i}^{m}\frac{(1-q^{3(m+i+j-1)})}{(1-q^{3(2i+j-1)})}\right\}. \tag{1.5}$$

If we let $m\to\infty$, then we obtain

Limiting Macdonald Conjecture.

$$\sum_{n\geqq 0} M(\infty,n)\,q^n=\prod_{n=1}^{\infty}\frac{(1-q^{3n-1})}{(1-q^{3n-2})(1-q^{3n})^{\left[\frac{n+1}{3}\right]}}. \tag{1.6}$$

The above conjectures are still open. Our main object in this paper is to prove the following:

Weak Macdonald Conjecture (Theorem 9 below). The total number of plane partitions π such that $D(\pi)\subseteq\mathscr{B}_{m,m,m}$ and $D(\pi)$ is G_3 invariant is

$$\prod_{i=1}^{m}\left\{\frac{3i-1}{3i-2}\prod_{j=i}^{m}\frac{m+i+j-1}{2i+j-1}\right\}$$

(i.e. (1.5) holds at $q=1$).

In our treatment of the Weak Macdonald Conjecture we are led to consider what we term descending plane partitions. To define these objects we must first consider shifted plane partitions, a topic treated extensively in [9] (see also [14] and [15]). A shifted plane partition ψ of n is an array of positive integers a_{ij} *defined only for* $j \geqq i$ such that $\sum a_{ij} = n$ (more briefly $|\psi| = n$), $a_{ij} \geqq a_{i,\,j+1}$, and for $j > i$, $a_{ij} \geqq a_{i+1,\,j}$. Thus the seven shifted plane partitions of 4 are

$$4,\quad 31,\quad 22,\quad 211,\quad \begin{array}{l}21,\\ \ 1\end{array}\quad 1111,\quad \begin{array}{l}111\\ \ 1\ .\end{array}$$

We shall call shifted plane partitions "strict" if there is strict decrease down each column. Finally we define a *descending plane partition* to be a strict shifted plane partition such that the first entry in each row does not exceed the number of parts in the preceding row but is larger than the number of parts in its own row. Thus

$$\begin{array}{cccc} 5 & 4 & 4 & 3 \\ & 3 & 3 & \\ & & 2 & \end{array}$$

is a descending plane partition of 24.

Descending Plane Partitions Conjecture. Let $De(m, n)$ denote the number of descending plane partitions of n in which each part is $\leqq m$. Then

$$\sum_{n \geqq 0} De(m, n)\, q^n = \prod_{i=1}^{m} \prod_{j=i}^{m} \frac{(1 - q^{m+i+j-1})}{(1 - q^{2i+j-1})}. \tag{1.7}$$

Limiting Descending Plane Partitions Conjecture.

$$\sum_{n \geqq 0} De(\infty, n)\, q^n = \prod_{n=1}^{\infty} \frac{1}{(1 - q^n)^{\left[\frac{n+1}{3}\right]}}. \tag{1.8}$$

These two conjectures are also open. However as a parallel consequence of our solution to the weak Macdonald conjecture we are able to prove also the following assertion:

Weak Descending Plane Partitions Conjecture (Theorem 10 below). The total number of descending plane partitions whose parts do not exceed m is $\prod_{i=1}^{m} \prod_{j=i}^{m} \frac{m+i+j-1}{2i+j-1}$.

In Sect. 2 we present a combinatorial study of shifted plane partitions with the object of showing their relationship to the conjectures we have described. The three main results to be proved there are:

Theorem 1. *Let* $\delta(a_1, \ldots, a_r; \lambda_1, \ldots, \lambda_r; n)$ denote the number of strict shifted r-rowed plane partitions of n with a_i the largest part in the i-th row and λ_i the number of parts in the i-th row. Then for $a_1 > a_2 > \ldots > a_r > 0$, $\lambda_1 > \lambda_2 > \ldots > \lambda_r > 0$,

$$\sum_{n\geq 0} \delta(a_1, \ldots, a_r; \lambda_1, \ldots, \lambda_r; n)\, q^n = q^{a_1+\ldots+a_r+\lambda_1+\ldots+\lambda_r-r} \det\left(\begin{bmatrix} a_j+\lambda_i-2 \\ \lambda_i-1 \end{bmatrix}\right), \tag{1.9}$$

where $\begin{bmatrix} A \\ B \end{bmatrix}$ is the Gaussian polynomial defined by

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{cases} \dfrac{(1-q^A)(1-q^{A-1})\ldots(1-q^{A-B+1})}{(1-q^B)(1-q^{B-1})\ldots(1-q)}, & B\geq 0 \\ 0, & B<0. \end{cases} \tag{1.10}$$

Theorem 3. *For* $m\geq 2$

$$\sum_{n\geq 0} De(m,n)\, q^n = \det\left(\delta_{ij} + q^{j+2}\begin{bmatrix} i+j+2 \\ i \end{bmatrix}\right)_{0\leq i,\, j\leq m-2}, \tag{1.11}$$

where $\delta_{ij}=0$ *if* $i\neq j$ *and* $\delta_{ii}=1$.

Theorem 4

$$\sum_{n\geq 0} M(m,n)\, q^n = \det\left(\delta_{ij} + q^{3j+1}\begin{bmatrix} i+j \\ i \end{bmatrix}_{q^3}\right)_{0\leq i,\, j\leq m-1} \tag{1.12}$$

where $\begin{bmatrix} A \\ B \end{bmatrix}_{q^3}$ *is the Gaussian polynomial with* q *replaced by* q^3.

Obviously then the Macdonald conjecture is reduced to evaluating the determinant in Theorem 4, while the descending plane partitions conjecture is reduced to evaluating the determinant in Theorem 3. Unfortunately we have not been able to evaluate these determinants except when $q=1$. In Sect. 3 we prove a hypergeometric series identity utilizing the work of Whipple [16, 17]. Then in Sect. 4, we prove the following result on determinants of binomial coefficients:

Theorem 8

$$\det\left(\delta_{ij} + \binom{\mu+i+j}{i}\right)_{0\leq i,\, j\leq m-1} = \prod_{j=0}^{m-1} \Delta_j(\mu), \tag{1.13}$$

where

$$\Delta_{2j}(\mu) = \begin{cases} \dfrac{(\mu+2j+2)_j \left(\dfrac{\mu}{2}+2j+\dfrac{3}{2}\right)_{j-1}}{(j)_j \left(\dfrac{\mu}{2}+j+\dfrac{3}{2}\right)_{j-1}}, & j>0 \\ 2, & j=0 \end{cases} \tag{1.14}$$

and

$$\Delta_{2j-1}(\mu) = \frac{(\mu+2j)_{j-1}\left(\dfrac{\mu}{2}+2j+\dfrac{1}{2}\right)_j}{(j)_j\left(\dfrac{\mu}{2}+j+\dfrac{1}{2}\right)_{j-1}}, \quad j>0, \tag{1.15}$$

with $(A)_j = A(A+1)\ldots(A+j-1)$.

It is then a straight forward exercise to show that the weak Macdonald conjecture (Theorem 9) follows from the case $\mu=0$ of Theorem 8, while the weak descending plane partitions conjecture follows from the case $\mu=2$.

One might expect that the proof of Theorem 8 would be easy especially in light of the fact that if the δ_{ij} is removed then it is easy to prove that

$$\det\left(\binom{\mu+i+j}{i}\right)_{0\leqq i,\, j\leqq m-1}=1, \quad [13;\text{p. }682].$$

However, it will become abundantly obvious in Sects. 3 and 4 that we have not found any simple proof of Theorem 8.

We conclude with an examination of the problems involved in proving the two main conjectures.

Since the results of this paper were first announced at Alfred Young Day in Waterloo on June 2, 1978, it is certainly fitting to include him in the dedication. Since the crucial results for the determinant evaluations (Sect. 3) rely heavily on the little known but highly significant work of Whipple [16, 17], we have also included him in the dedication.

2. Shifted and Descending Plane Partitions

The cornerstone of all our work is Theorem 1 (stated in Sect. 1) which allows us to represent various generating functions as determinants. Our proof resembles that of Carlitz [8] for his derivation of (1.2).

Before we treat Theorem 1 we must make some conventions which vary from the ordinary theory of partitions. Namely, we shall admit into consideration two different partitions of zero: (i) the "empty" partition of zero (i.e. that partition that has no parts) will be said to have no parts and largest part equal to zero; (ii) the "non-empty" partition of zero is that partition which has one part and that part is zero. These rather strange conventions greatly simplify the recurrences that we must treat.

Proof of Theorem 1. When $r=1$, $\delta(a_1;\lambda_1;n)$ is the number of partitions of n into exactly λ_1 parts with largest part equal to a_1. Hence by Theorem 3.1 of [3; p. 33] we find

$$\begin{aligned}
&\sum_{n\geqq 0}\delta(a_1;\lambda_1;n)\,q^n\\
&=\begin{cases} q^{a_1+\lambda_1-1}\begin{bmatrix}(a_1-1)+(\lambda_1-1)\\ \lambda_1-1\end{bmatrix}, & \text{if } a_1\geqq 0,\ \lambda_1\geqq 0,\ \textit{not both}=0\\ 1, & a_1=\lambda_1=0.\end{cases}\\
&=\begin{cases} q^{a_1+\lambda_1-1}\begin{bmatrix}a_1+\lambda_1-2\\ \lambda_1-1\end{bmatrix}, & \text{if } a_1\geqq 0,\ \lambda_1\geqq 0,\ \text{not both}=0\\ 1, & \text{if } a_1=\lambda_1=0.\end{cases}
\end{aligned}\tag{2.1}$$

We next define

$$G(a_1,\ldots,a_r;\lambda_1,\ldots,\lambda_r;q)=\sum_{n\geqq 0}\delta(a_1,\ldots,a_r;\lambda_1,\ldots,\lambda_r;n)\,q^n. \tag{2.2}$$

Instead of proving (1.9) for $a_1 > a_2 > \ldots > a_r \geqq 1$, $\lambda_1 > \ldots > \lambda_r \geqq 1$, we prove more generally that for $a_1 > \ldots > a_r \geqq 0$, $\lambda_1 > \ldots > \lambda_r \geqq 0$,

$$G(a_1, \ldots, a_r; \lambda_1, \ldots, \lambda_r; q) = \det(G(a_i; \lambda_j; q))_{1 \leqq i, j \leqq r}. \tag{2.3}$$

We proceed by a double induction first on r then on λ_r. If $r=1$, (2.3) is trivial and (2.1) provides the Gaussian polynomial representation of $G(a_i; \lambda_1; q)$.

Assume that (2.3) is valid up to but not including a fixed r. We now start a second induction on λ_r. If $\lambda_r = 0$, we see that

$$\begin{aligned} &G(a_1, \ldots, a_r; \lambda_1, \ldots, \lambda_{r-1}, 0; q) \\ &\quad = \begin{cases} G(a_1, \ldots, a_{r-1}; \lambda_1, \ldots, \lambda_{r-1}; q), & \text{if } a_r = 0 \\ 0, & \text{if } a_r > 0. \end{cases} \end{aligned} \tag{2.4}$$

However expanding the determinant in (2.3) along the right hand column we see that with $\lambda_r = 0$:

$$\det(G(a_i; \lambda_j; q))_{1 \leqq i, j \leqq r} = G(a_r; 0; q) \det(G(a_i; \lambda_j; q)_{1 \leqq i, j \leqq r}, \tag{2.5}$$

and from (2.1) we see that (2.5) is the same assertion for the right side of (2.3) as (2.4) is for the left side. Consequently by the induction hypothesis on r, (2.3) holds for $\lambda_r = 0$. We now assume $\lambda_r > 0$; consequently a_r must be >0. Hence for $\lambda_j \geqq 0$, we have (assuming $a_{r+1} = -1$)

$$\sum_{b_i = a_{i+1}+1}^{a_i} G(b_i; \lambda_j - 1; q) = q^{-a_i} G(a_i; \lambda_j; q) - q^{-a_{i+1}} G(a_{i+1}; \lambda_j; q); \tag{2.6}$$

this identity is the assertion for $\lambda_j > 1$ that (by [3; p. 37, Eq. (3.3.9))

$$\sum_{b_i = a_{i+1}+1}^{a_i} q^{b_i + \lambda_j - 2} \begin{bmatrix} b_i + \lambda_j - 3 \\ \lambda_j - 2 \end{bmatrix} = q^{\lambda_j - 1} \left(\begin{bmatrix} a_i + \lambda_j - 2 \\ \lambda_{j+1} - 1 \end{bmatrix} - \begin{bmatrix} a_{i+1} + \lambda_j - 2 \\ \lambda_j - 1 \end{bmatrix} \right), \tag{2.7}$$

and for $\lambda_j = 1$ (and therefore $j = r$) that $0 = q^{-a_i} \cdot q^{a_i} - q^{-a_{i+1}} \cdot q^{a_{i+1}}$ if $i < r$, and $1 = q^{-a_r} q^{a_r}$.

By directing attention to what remains if we remove each of the first parts in each row of a strict shifted plane partition, we see that (where $a_{r+1} = -1$)

$$\begin{aligned} &G(a_1, \ldots, a_r; \lambda_1, \ldots, \lambda_r; q) \\ &\quad = q^{a_1 + \ldots + a_r} \sum_{\substack{a_1 \geqq b_1 > a_2 \\ a_2 \geqq b_2 > a_3 \\ \vdots \\ a_r \geqq b_r > a_r + 1}} G(b_1, \ldots, b_r; \lambda_1 - 1, \ldots, \lambda_r - 1; q). \end{aligned} \tag{2.8}$$

However it is true that the right side of (2.3) also satisfies this recurrence, a fact which may be seen as follows:

$$\begin{aligned}
&q^{a_1+\ldots+a_r} \sum_{\substack{a_1 \geqq b_1 > a_2 \\ a_2 \geqq b_2 > a_3 \\ \vdots \\ a_r \geqq b_r > a_{r+1}}} \det(G(b_i; \lambda_j - 1; q)) \\
&= q^{a_1+\ldots+a_r} \det(q^{-a_i} G(a_i; \lambda_j; q) - q^{-a_{i+1}} G(a_{i+1}; \lambda_j; q)) \\
&\qquad \text{(applying (2.6) to each row)} \\
&= q^{a_1+\ldots+a_r} \det(q^{-a_i} G(a_i; \lambda_j; q)) \\
&\qquad \text{(noting } G(a_{r+1}; \lambda_j; q) = 0 \text{ since } \lambda_j \geqq \lambda_r > 0\text{, and starting} \\
&\qquad \text{from the bottom adding each row to the one above it)} \\
&= \det(G(a_i; \lambda_j; q)). \qquad (2.9)
\end{aligned}$$

Therefore by the fact that each side of (2.3) satisfies the recurrence (2.8) we see that therefore (2.3) holds for every λ_r by the induction on λ_r and therefore for every r by the induction on r. □

The following theorem provides a further application of Carlitz's technique of [8]. It allows us to treat several families of plane partition problems.

Theorem 2. *For $d \leqq 2$, and fixed $a_1 > a_2 > \ldots > a_r \geqq 0$*

$$\sum_{a_1+d > \lambda_1 \geqq a_2+d > \ldots \geqq a_r+d > \lambda_r \geqq 1} \det(G(a_i; \lambda_j; q)) = q^{a_1+\ldots+a_r} \det\left(\begin{bmatrix} a_i + a_j - 2 + d \\ a_j \end{bmatrix}\right) \tag{2.10}$$

where the λ_i are the summation indices.

Proof. In parallel with (2.6), we now require (assuming $a_{r+1} = -d+1$)

$$\sum_{\lambda_j = a_{j+1}+d}^{a_j+d-1} G(a_i; \lambda_j; q) = q^{a_i}\left(\begin{bmatrix} a_i + a_j + d - 2 \\ a_i \end{bmatrix} - \begin{bmatrix} a_i + a_{j+1} + d - 2 \\ a_i \end{bmatrix}\right); \tag{2.11}$$

this result is equivalent to

$$\sum_{\lambda_j = a_{j+1}+d}^{a_j+d-1} q^{a_i+\lambda_j-1} \begin{bmatrix} a_i + \lambda_j - 2 \\ \lambda_j - 1 \end{bmatrix} = q^{a_i}\left(\begin{bmatrix} a_i + a_j + d - 2 \\ a_i \end{bmatrix} - \begin{bmatrix} a_i + a_{j+1} + d - 2 \\ i \end{bmatrix}\right). \tag{2.12}$$

Thus applying (2.11) to each column independently in (2.10), we find that the left side of (2.10) equals

$$\begin{aligned}
&q^{a_1+\ldots+a_r} \det\left(\begin{bmatrix} a_i + a_j - 2 + d \\ a_i \end{bmatrix} - \begin{bmatrix} a_i + a_{j+1} - 2 + d \\ a_i \end{bmatrix}\right) \\
&= q^{a_1+\ldots+a_r} \det\left(\begin{bmatrix} a_i + a_j - 2 + d \\ a_j \end{bmatrix}\right),
\end{aligned}$$

since $\begin{bmatrix} a_i + a_{r+1} - 2 + d \\ a_i \end{bmatrix} = \begin{bmatrix} a_i - 1 \\ a_i \end{bmatrix} = 0$ the second determinant is obtained from the first by adding each column to the one on its left starting from the right. □

We are now prepared to prove a theorem more general than Theorem 3 (stated in the Introduction). We let $De(d; m, n)$ denote the number of strict shifted plane partitions of n with largest part $\leqq m+1-d$ such that the largest part on each row is $\leqq$ the number of parts on the previous row less d at the same time being larger than the number of parts on the same row less d. Note that $De(0; m, n) = De(m+1, n)$.

Theorem 3′. *For $d \leqq 2$ and integral,*

$$\sum_{n \geqq 0} De(d; m, n)\, q^n = \det\left(\delta_{ij} + q^{j-d+2}\begin{bmatrix} i+j+2-d \\ i \end{bmatrix}\right)_{0 \leqq i,\, j \leqq m-1}. \tag{2.13}$$

Proof. From (2.3) and Theorem 2, we see that

$$\sum_{n \geqq 0} De(d; m, n)\, q^n = \sum_{\substack{1+m-d \geqq a_1 > a_2 > \ldots > a_r \geqq 2-d \\ r \text{ arbitrary}}} \det\left(q^{a_j}\begin{bmatrix} a_i + a_j - 2 + d \\ a_j \end{bmatrix}\right).$$

But this is merely the formula for the expansion along the main diagonal of

$$\det\left(\delta_{ij} + q^j\begin{bmatrix} i+j-2+d \\ j \end{bmatrix}\right)_{2-d \leqq i,\, j \leqq 1+m-d}$$
$$= \det\left(\delta_{ij} + q^{j-d+2}\begin{bmatrix} i+j-d+2 \\ i \end{bmatrix}\right)_{0 \leqq i,\, j \leqq m-1}.$$

We next utilize Theorem 2 to prove Theorem 4 (stated in the Introduction).

Proof of Theorem 4. We begin with an arbitrary plane partition π that is invariant under G_3. We form a new semi-graphical representation for π as follows:

(1) Associate a 1 with each point (i, i, i) in the positive 1st octant that lies in $D(\pi)$.

(2) Associate a 3 with each point (i, j, k) in the positive first octant that lies in $D(\pi)$ for which $k < \min(j, i)$ or $k = j < i$.

This process clearly associates with each π invariant under G_3 a set of two dimensional arrays (for each $k \geqq 1$). For example the arrays

$$\begin{array}{lllll} k=1 & & & & \\ 1 & 3 & 3 & 3 & 3 \\ & 3 & 3 & 3 & \\ & 3 & 3 & & \end{array} \qquad \begin{array}{ll} k=2 & \\ 1 & 3 \end{array}$$

correspond to

$$\begin{array}{lllll} 5 & 4 & 3 & 1 & 1 \\ 3 & 3 & 2 & 1 & \\ 3 & 2 & 1 & & \\ 2 & & & & \\ 1 & & & & \end{array}.$$

Notice that the 1's and 3's are associated with corresponding representatives of orbits of size 1 and 3 respectively. Consequently the sum total of the 1's and 3's in all the layers equals $|\pi|$. Now we add up the entries in each row of each layer and write the results as a shifted plane partition. Thus the above example yields

$$\begin{matrix} 13 & 9 & 6 \\ & 4 & . \end{matrix}$$

note that the first entries in each row are of the form $3a_{ii}+1=3a_i+1$, while the remaining entries are of the form $3a_{ij}$. Furthermore in order that the shifted plane partition π' come from a plane partition π invariant under G_3 it is clearly necessary and sufficient that

$$a_1>\lambda_1-2\geqq a_2>\lambda_2-2\geqq\ldots\geqq a_r>\lambda_r-2\geqq -1,$$

where λ_i is the number of parts on the i-th row of π'. Note that $a_r=0$ is permissable since it corresponds to the r-th layer having the single entry 1. Applying (2.3) and Theorem 2, we thus see that

$$\begin{aligned}
&\sum_{n\geqq 0} M(m,n)\,q^n\\
&\quad=\sum_{m+1\geqq a_1+2>\lambda_1\geqq a_2+2>\ldots>a_r+2>\lambda_r\geqq 1} q^r\, G(a_1,\ldots,a_r;\lambda_1,\ldots,\lambda_r;q^3)\\
&\quad=\sum_{m+1\geqq a_1+2>\lambda_1\geqq a_2+2>\ldots\geqq a_r+2>\lambda_r\geqq 1} q^r \det(G(a_i;\lambda_j;q^3))\\
&\quad=\sum_{m-1\geqq a_1>a_2>\ldots>a_r\geqq 0} \det\left(q^{3a_i+1}\begin{bmatrix} a_i+a_j\\ a_j\end{bmatrix}_{q^3}\right)\\
&\quad=\det\left(\delta_{ij}+q^{3j+1}\begin{bmatrix} i+j\\ i\end{bmatrix}_{q^3}\right)_{0\leqq i,\,j\leqq m-1}.
\end{aligned}$$

3. Hypergeometric Series

In order to evaluate the determinant in Theorem 8, we require the following hypergeometric series identity which follows from some fundamental transformations due to F.J.W. Whipple [16, 17].

Theorem 5. *Let*

$$M(i,j,m;a,\omega)=\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}} \tag{3.1}$$

$$\times\sum_{s\geqq 0}\binom{m+i+j+a-s-1}{i-j+a+s-1}\frac{(j-s)_{2s}(-m-3j-a-\omega+1)_s 4^{-s}}{s!\left(-\frac{m}{2}-2j-a-\omega+2\right)_s\left(-j-\frac{m}{2}+\frac{1}{2}\right)_s},$$

and

$$A(i,j,m;a,\omega)=\frac{(2j-a-i)!}{(2j+\omega-i-1)!}(-1)^{i+a}\sum_{s\geqq 0}\binom{j-s-1}{i-j+a+s-1}$$
$$\times\frac{(j-s)_{2s+a+\omega-1}}{s!\,(m+4j-2)(m+4j-4)\dots(m+4j-2s)}$$
$$\times\frac{(-m-3j-a-\omega+1)_s}{(m+2j-1)(m+2j-3)\dots(m+2j-2s+1)}. \tag{3.2}$$

Then for a,ω,i and j integers with $a+\omega\geqq 1$, $2j-a\geqq i$,

$$M(i,j,m;\,a,\omega)=A(i,j,m;\,a,\omega). \tag{3.3}$$

Remark. For a restatement of Theorem 5 as a purely hypergeometric identitiy see Sect. 5, Eq. (5.1).

Proof. We shall require formulas involving the generalized hypergeometric function:

$${}_{s+1}F_s\left[\begin{matrix}a_0,a_1,\dots,a_s;\,t\\ b_1,\dots,b_s\end{matrix}\right]=\sum_{n\geqq 0}\frac{(a_0)_n(a_1)_n\dots(a_s)_n\,t^n}{n!\,(b_1)_n\dots(b_s)_n}, \tag{3.4}$$

where

$$(A)_n=A(A+1)\dots(A+n-1)=\frac{\Gamma(A+n)}{\Gamma(A)},$$

[7; p. 8].

The two formulas we need are as follows (M and N nonnegative integers):

$${}_4F_3\left[\begin{matrix}A,B,C,-M;\,1\\ R,S,A+B+C-R-S+1-M\end{matrix}\right]$$
$$=\frac{(R+S-A-B)_M(S-C)_M}{(R+S-A-B-C)_M(S)_M}\times{}_4F_3\left[\begin{matrix}C,R-B,R-A,-M;\,1\\ R,R+S-A-B,\,C-S+1-M\end{matrix}\right],$$
$$[16;\text{ Eq. }(10.11)], \tag{3.5}$$

$${}_4F_3\left[\begin{matrix}-N,B,C,D;\,1\\ 1-N-B,1-N-C,W\end{matrix}\right]$$
$$=\frac{(W-D)_N}{(W)_N}\,{}_5F_4\left[\begin{matrix}D,1-N-B-C,-\frac{N}{2},\frac{1}{2}-\frac{N}{2},1-N-W;\,1\\ 1-N-B,1-N-C,\frac{1}{2}(1+D-W-N),1+\frac{1}{2}(D-W-N)\end{matrix}\right]$$
$$[17;\text{ Eq. }(6.6)],\ [7;\text{ p. }33,\text{ Eq. }(1)]. \tag{3.6}$$

Identity (3.5) is Whipple's relation between two Saalschutzian or balanced ${}_4F_3$'s; the hypergeometric series given by (3.4) is called Saalschutzian or balanced if $1+a_0+a_1+\dots+a_s=b_1+b_2+\dots+b_s$. It should be noted in passing that (3.5) is a little known generalization of the celebrated Pfaff-Saalschutz summation [7; p. 9] which follows from (3.5) if we set $R=A$. Identity (3.6) is a transformation of a nearly poised ${}_4F_3$ series of the second kind into a balanced or Saalschutzian ${}_5F_4$; the hypergeometric series given by (3.4) is called "nearly poised of the second kind" if $1+a_0=a_1+b_1=a_2+b_2=\ldots=a_{s-1}+b_{s-1}$.

$$M(i,j,m;\,a,\omega)=\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}}$$

$$\times\sum_{s\geq 0}\binom{m+i+j+a-s-1}{i-j+a+s-1}\frac{(j-s)_{2s}(-m-3j-a-\omega+1)_s\cdot 4^{-s}}{s!\left(-\frac{m}{2}-2j-a-\omega+2\right)_s\left(-j-\frac{m}{2}+\frac{1}{2}\right)_s}$$

$$=\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}}\sum_{s\geq 0}\frac{(m+i+j+a-1)!\,(j-s)_{2s}(-m-3j-a-\omega+1)_s\,2^{-s}}{(i-j+a-1)!\,(m+2j-2s)!\left(-\frac{m}{2}-2j-a-\omega+2\right)_s}$$

$$\times\frac{1}{(-m-i-j-a+1)_s(i-j+a)_s(m+2j-1)(m+2j-3)\ldots(m+2j-2s+1)}$$

$$=\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}}\binom{m+i+j+a-1}{i-j+a-1}$$

$$\times{}_4F_3\left[\begin{matrix}-j+1,j,\;-m-3j-a-\omega+1,\;-\frac{m}{2}-j;\;1\\ i-j+a,\;-\frac{m}{2}-2j-a-\omega+2,\;-m-i-j-a+1\end{matrix}\right]$$

$$=\binom{m+i+j+a-1}{i-j+a-1}\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}}$$

$$\times\frac{\left(i-j+\frac{m}{2}+a+1\right)_{j-1}(-j-a-\omega+2)_{j-1}}{(i+m+a+1)_{j-1}\left(-\frac{m}{2}-2j-a-\omega+2\right)_{j-1}}$$

$$\times{}_4F_3\left[\begin{matrix}i-2j+a,\;-j+1,\;-j-\frac{m}{2},\;m+i+2j+2a+\omega-1;\;1\\ i-j+a,\;i-j+\frac{m}{a}+a+1,\;a+\omega\end{matrix}\right]$$

(by (3.5) with $A=j$, $B=-m-3j-a-\omega+1$,

$$C=-j-\frac{m}{2}, R=i-j+a, S=-\frac{m}{2}-2j-a-\omega+2\Big)$$

$$=\binom{m+i+j+a-1}{i-j+a-1}\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}}$$

$$\times\frac{\left(i-j+\frac{m}{2}+a+1\right)_{j-1}(-j-a-\omega+2)_{j-1}}{(i+m+a+1)_{j-1}\left(-\frac{m}{2}-2j-a-\omega+2\right)_{j-1}}\frac{(-m-i-2j-a+1)_{2j-i-a}}{(a+\omega)_{2j-i-a}}$$

$$\times{}_4F_3\left[\begin{matrix}m+i+2j+2a+\omega-1,\ -j+\frac{i+a}{2},\ -j+\frac{i+a+1}{2},\ 1+i-2j-\omega;\ 1\\ i-j+a,\ i-j+\frac{m}{2}+a+1,\ 1+\frac{1}{2}(m+2i+2a-1)\end{matrix}\right]$$

$\Big($by (3.6) with $N=2j-i-a$, $B=-j+1$, $C=-j-\frac{m}{2}$,

$$D=m+i+2j+2a+\omega-1,\ W=a+\omega\Big)$$

$$=\binom{m+i+j+a-1}{i-j+a-1}\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}\left(i-j+\frac{m}{2}+a+1\right)_{j-1}(-j-a-\omega+2)_{j-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}(i+m+a+1)_{j-1}\left(-\frac{m}{2}-2j-a-\omega+2\right)_{j-1}}$$

$$\times\frac{(-m-i-2j-a+1)_{2j-i-a}\,\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+\frac{1}{2}\right)\Gamma\left(i-j+\frac{m}{2}+a+1\right)}{(a+\omega)_{2j-i-a}\,\Gamma\left(\frac{i+a+1+m}{2}\right)\Gamma\left(\frac{i+a+2+m}{2}\right)}$$

$$\times\frac{\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+1\right)\Gamma\left(\frac{m}{2}+j+\frac{1}{2}\right)}{\Gamma\left(1-\frac{m}{2}-2j\right)\Gamma\left(-\frac{m}{2}-i-a+\frac{1}{2}\right)}$$

$$\times{}_4F_3\left[\begin{matrix}-j+\frac{i+a}{2},\ -j+\frac{i+a+1}{2},\ j+a+\omega-1,\ -m-3j-a-\omega+1;1\\ i-j+a,\ -\frac{m}{2}-2j+1,\ -\frac{m}{2}-j+\frac{1}{2}.\end{matrix}\right]$$

$\Big($by (3.5) once we have noted that it is symmetric in C and $-M$ so that we need only require one of them to be a nonpositive integer; we therefore may take A

$=1+i-2j-\omega,\ B=m+i+2j+2a+\omega-1,\ C=-j+\dfrac{i+a+1}{2},\quad -M=-j+\dfrac{i+a}{2},$
$R=i-j+a,\ S=i-j+\dfrac{m}{2}+a+1\Big).$

The ${}_4F_3$ series that we now have is basically what we want for $\Lambda(i,j,m;a,\omega)$. First we must simplify the massive expression in front of the ${}_4F_3$. In our treatment we shall assume that m is not an integer; the case when m is an integer (i.e. the interesting case for us) will then be obtained by continuity. We shall utilize (3.5) as well as Legendre's duplication formula [18; p. 240]:

$$\Gamma(2x)=\frac{2^{2x-1}}{\sqrt{\pi}}\,\Gamma(x)\,\Gamma(x+\tfrac{1}{2}), \tag{3.7}$$

and the functional equation [18; p. 239]:

$$\Gamma(x)\,\Gamma(1-x)=\frac{\pi}{\sin \pi x}. \tag{3.8}$$

$$\binom{m+i+j+a-1}{i-j+a-1}\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}\left(i-j+\frac{m}{2}+a+1\right)_{j-1}(-j-a-\omega+2)_{j-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}(i+m+a+1)_{j-1}\left(-\frac{m}{2}-2j-a-\omega+2\right)_{j-1}}$$

$$\times\frac{(-m-i-2j-a+1)_{2j-i-a}\,\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+\frac{1}{2}\right)\Gamma\left(i-j+\frac{m}{2}+a+1\right)}{(a+\omega)_{2j-i-a}\,\Gamma\left(\frac{i+a+1+m}{2}\right)\Gamma\left(\frac{i+a+2+m}{2}\right)}$$

$$\times\frac{\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+1\right)\Gamma\left(\frac{m}{2}+j+\frac{1}{2}\right)}{\Gamma\left(1-\frac{m}{2}-2j\right)\Gamma\left(-\frac{m}{2}-i-a+\frac{1}{2}\right)}$$

$$=\binom{m+i+j+a-1}{i-j+a-1}\frac{\left(i-j+\frac{m}{2}+a+1\right)_{j-1}(a+\omega)_{j-1}}{(i+m+a+1)_{j-1}\left(\frac{m}{2}+j+a+\omega\right)_{j-1}}$$

$$\times\frac{(-1)^{i+a}(m+2i+2a)_{2j-i-a}(a+\omega-1)!}{(2j-i+\omega-1)!}$$

$$\times\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+\frac{1}{2}\right)}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}\Gamma\left(\frac{i+a+1+m}{2}\right)\Gamma\left(\frac{i+a+2+m}{2}\right)}$$

$$\times\frac{\Gamma\left(i-j+\frac{m}{2}+a+1\right)\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+1\right)\Gamma\left(\frac{m}{2}+j+\frac{1}{2}\right)}{\Gamma\left(1-\frac{m}{2}-2j\right)\Gamma\left(-\frac{m}{2}-i-a+\frac{1}{2}\right)}$$

$$=\frac{(-1)^{i+a}\Gamma(m+i+a+1)\left(i-j+\frac{m}{2}+a+1\right)_{j-1}}{(i-j+a-1)!\,\Gamma(m+2j+1)\left(\frac{m}{2}+j+a+\omega\right)_{j-1}}$$

$$\times\frac{(a+\omega+j-2)!\,(m+2i+2a)_{2j-i-a}}{(2j-i+\omega-1)!}$$

$$\times\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+\frac{1}{2}\right)\Gamma\left(i-j+\frac{m}{2}+a+1\right)}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}\Gamma\left(\frac{i+a+1+m}{2}\right)\Gamma\left(\frac{i+a+2+m}{2}\right)}$$

$$\times\frac{\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+1\right)\Gamma\left(\frac{m}{2}+j+\frac{1}{2}\right)}{\Gamma\left(1-\frac{m}{2}-2j\right)\Gamma\left(-\frac{m}{2}-i-a+\frac{1}{2}\right)}$$

$$=(-1)^{i+a}\binom{j+a+\omega-2}{i-j+a-1}\frac{\Gamma(i+a+m+1)}{\Gamma(m+2j+1)}$$

$$\times\frac{\left(i-j+\frac{m}{2}+a+1\right)_{j-1}(m+2i+2a)_{2j-i-a}}{\left(\frac{m}{2}+j+a+\omega\right)_{j-1}}$$

$$\times\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+\frac{1}{2}\right)\Gamma\left(i-j+\frac{m}{2}+a+1\right)}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}\Gamma\left(\frac{i+a+1+m}{2}\right)\Gamma\left(\frac{i+a+2+m}{2}\right)}$$

$$\times\frac{\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+1\right)\Gamma\left(\frac{m}{2}+j+\frac{1}{2}\right)}{\Gamma\left(1-\frac{m}{2}-2j\right)\Gamma\left(-\frac{m}{2}-i-a+\frac{1}{2}\right)}$$

$$=(-1)^{i+a}\binom{j+a+\omega-2}{i-j+a-1}\frac{\Gamma(m+i+a+1)\Gamma\left(i+\frac{m}{2}+a\right)\Gamma\left(\frac{m}{2}+j+a+\omega\right)}{\Gamma(m+2j+1)\Gamma\left(\frac{m}{2}+2j+a+\omega-1\right)\Gamma(m+2i+2a)}$$

$$\times\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}\Gamma(m+2j+i+a)\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+\frac{1}{2}\right)}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}\Gamma\left(\frac{i+a+1+m}{2}\right)\Gamma\left(\frac{i+a+2+m}{2}\right)}$$

$$\times\frac{\Gamma\left(-\frac{m}{2}-j-\frac{i+a}{2}+1\right)\Gamma\left(\frac{m}{2}+j+\frac{1}{2}\right)}{\Gamma\left(1-\frac{m}{2}-2j\right)\Gamma\left(-\frac{m}{2}-i-a+\frac{1}{2}\right)}$$

$$=(-1)^{i+a}\binom{j+a+\omega-2}{i-j+a-1}2^{2(i+a+m+j)-1}\Gamma(-m-2j-i-a+1)$$

$$\times\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}\Gamma\left(i+\frac{m}{2}+a\right)\Gamma\left(\frac{m}{2}+j+a+\omega\right)}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}\Gamma(m+2j+1)\Gamma\left(\frac{m}{2}+2j+a+\omega-1\right)}$$

$$\times\frac{\Gamma(m+2j+i+a)\Gamma\left(\frac{m}{2}+j+\frac{1}{2}\right)}{\Gamma(m+2i+a)\Gamma\left(1-\frac{m}{2}-2j\right)\Gamma\left(-\frac{m}{2}-i-a+\frac{1}{2}\right)}$$

$$=(-1)^{i+a}\binom{j+a+\omega-2}{i-j+a-1}\frac{2\pi\sin\pi\left(i+\frac{m}{2}+a+\frac{1}{2}\right)}{\sin\pi(m+2j+i+a)}$$

$$\times\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}\Gamma\left(\frac{m}{2}+j+a+\omega\right)}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}\Gamma\left(\frac{m}{2}+j+1\right)\Gamma\left(\frac{m}{2}+2j+a+\omega-1\right)\Gamma\left(1-\frac{m}{2}-2j\right)}$$

$$=(-1)^{i+a}\binom{j+a+\omega-2}{i-j+a-1}\frac{2\sin\pi\left(i+\frac{m}{2}+a+\frac{1}{2}\right)\sin\frac{\pi m}{2}}{\sin\pi(m+i+a)}$$

$$=(-1)^{i+a}\binom{j+a+\omega-2}{i-j+a-1}\frac{2(-1)^{i+a}\cos\frac{\pi m}{2}\sin\frac{\pi m}{2}}{(-1)^{i+a}\sin\pi m}$$

$$=(-1)^{i+a}\binom{j+a+\omega-2}{i-j+a-1}\quad(\text{since }\sin 2\theta=2\sin\theta\cos\theta).$$

Hence we have

$$M(i,j,m;\,a,\omega)=(-1)^{i+a}\binom{j+a+\omega-2}{i-j+a-1}$$

$$\times{}_4F_3\left[\begin{matrix}-j+\frac{i+a}{2},\ -j+\frac{i+a+1}{2},\ j+a+\omega-1,\ -m-3j-a-\omega+1;\ 1\\ i-j+a,\ -\frac{m}{2}-2j+1,\ -\frac{m}{2}-j+\frac{1}{2}\end{matrix}\right]$$

$$=(-1)^{i+a}\sum_{s\geqq 0}\frac{(j+a+\omega+s-2)!(-2j+i+a)_{2s}}{s!(i-j+a+s-1)!(2j-i+\omega-1)!}$$

$$\times\frac{4^{-s}(-m-3j-a-\omega+1)_s}{\left(-\frac{m}{2}-2j+1\right)_s\left(-\frac{m}{2}-j+\frac{1}{2}\right)_s}=\frac{(-1)^{i+a}(2j-i-a)!}{(2j-i+\omega-1)!}$$

$$\times \sum_{s \geq 0} \binom{j-s-1}{i-j+a+s-1} \frac{4^{-s}(-m-3j-a-\omega+1)_s}{s!\left(-\frac{m}{2}-2j+1\right)_s \left(-\frac{m}{2}-j+\frac{1}{2}\right)_s}$$

$$\times \frac{(j+a+\omega+s-2)!}{(j-s-1)!}$$

$$= \frac{(-1)^{i+a}(2j-i-a)!}{(2j-i+\omega-1)!} \sum_{s \geq 0} \binom{j-s-1}{i-j+a+s-1} \frac{(-m-3j-a-\omega+1)_s}{s!}$$

$$\times \frac{(j-s)_{2s+a+\omega-1}}{(m+4j-2)(m+4j-4)\ldots(m+4j-2s)(m+2j-1)(m+2j-3)\ldots(m+2j-2s+1)}$$

$= \Lambda(i, j, m;\, a, \omega)$, as desired.

In actual fact our above treatment would appear to be valid only provided $i-2j+a$ is a nonpositive integer and $i-j+a$ is any complex (or real) number other than a nonpositive integer. The result then follows for $i-j+a$ nonpositive integral by continuity. Also we had to assume m nonintegral to allow $\sin \pi m$ in a denominator; however again the excluded values on m are admissible by continuity. □

Corollary 5a

$$\Lambda(i, j, m;\, a, \omega) = \begin{cases} \binom{j+a+\omega-2}{j-1} & \textit{for } i = 2j-a \\ 0 & \textit{for } i > 2j-a. \end{cases}$$

Proof. In (3.2) we see that if $i = 2j-a$ then only the first term does not vanish while if $i > 2j-a$ then all terms vanish. □

Corollary 5b. *For j a positive integer,*

$$M(2j+\omega, j, m;\, a, \omega)$$
$$= \binom{m+2j+a+\omega-1}{a+\omega-1} \frac{(m+2j+a+\omega)_j \left(\frac{m}{2}+2j\right)_{a+\omega-1} \left(2j+\frac{m}{2}+a+\omega\right)_{j-1}}{(j+a+\omega-1)_j \left(\frac{m}{2}+j+1\right)_{a+\omega-1} \left(\frac{m}{2}+j+a+\omega\right)_{j-1}}.$$

Proof. After the first transformation applied to $M(i, j, m;\, a, \omega)$ in the proof of Theorem 5 we find that

$$M(2j+\omega, j, m;\, a, \omega)$$
$$= \binom{m+3j+a+\omega-1}{j+a+\omega-1} \frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1} \left(j+\frac{m}{2}+a+\omega+1\right)_{j-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1} (2j+m+a+\omega+1)_{j-1}}$$

$$\times \frac{(-j-a-\omega+2)_{j-1}}{\left(-\frac{m}{2}-2j-a-\omega+2\right)_{j-1}}\, {}_3F_2 \left[\begin{matrix} -j+1,\ -j-\frac{m}{2},\ m+4j+2a+2\omega-1;\ 1 \\ j+a+\omega,\ j+\frac{m}{2}+a+\omega+1 \end{matrix}\right]$$

$$=\binom{m+3j+a+\omega-1}{j+a+\omega-1}\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}\left(j+\frac{m}{2}+a+\omega+1\right)_{j-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}(2j+m+a+\omega+1)_{j-1}}$$

$$\times\frac{(a+\omega)_{j-1}\left(2j+\frac{m}{2}+a+\omega\right)_{j-1}(m+2j+a+\omega+1)_{j-1}}{\left(\frac{m}{2}+j+a+\omega\right)_{j-1}(j+a+\omega)_{j-1}\left(\frac{m}{2}+j+a+\omega+1\right)_{j-1}}\qquad\text{(by [7; p. 9])}$$

$$=\frac{(m+2j+a+\omega-1)!\,(m+2j+a+\omega)_j}{(a+\omega-1)!\,(j+a+\omega-1)(m+2j)!}$$

$$\times\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}\left(2j+\frac{m}{2}+a+\omega\right)_{j-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}\left(\frac{m}{2}+j+a+\omega\right)_{j-1}(j+a+\omega)_{j-1}}$$

$$=\binom{m+2j+a+\omega-1}{a+\omega-1}\frac{(m+2j+a+\omega)_j\left(\frac{m}{2}+2j\right)_{a+\omega-1}\left(2j+\frac{m}{2}+a+\omega\right)_{j-1}}{(j+a+\omega-1)_j\left(\frac{m}{2}+j+1\right)_{a+\omega-1}\left(\frac{m}{2}+j+a+\omega\right)_{j-1}},$$

as desired. □

While Theorem 5 and its corollaries play the central role in the next section, it turns out that the following relation among three balanced ${}_4F_s$'s is crucial in order to conclude the proof of Theorem 7.

Theorem 6. *For $0\leqq i\leqq 2j-2$ with $i-2j+2$ integral,*

$$\begin{aligned}&{}_4F_3\left[\begin{matrix}i-2j+1,\,1-j,\,-j-\frac{\mu}{2}+\frac{1}{2},\,i+\mu+2j;\,1\\ i-j+1,\,i-j+\frac{\mu}{2}+\frac{3}{2},\,1\end{matrix}\right]+\frac{(2i+\mu+1)\left(\frac{\mu}{2}+j-\frac{1}{2}\right)j(j-1)}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)(i-j+1)(i-j+2)}\\
&\times{}_4F_3\left[\begin{matrix}i-2j+2,\,2-j,\,-j-\frac{\mu}{2}+\frac{3}{2},\,i+\mu+2j+1;\,1\\ i-j+3,\,i-j+\frac{\mu}{2}+\frac{5}{2},\,2\end{matrix}\right]\\
&+\frac{\left(\frac{\mu}{2}+i+\frac{1}{2}\right)(i+1)(2j-i-2)j}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)(i-j+1)(i-j+2)}\\
&\times{}_4F_3\left[\begin{matrix}i-2j+3,\,1-j,\,-j-\frac{\mu}{2}+\frac{3}{2},\,i+\mu+2j+1;\,1\\ i-j+3,\,i-j+\frac{\mu}{2}+\frac{5}{2},\,2\end{matrix}\right]=0.\end{aligned}\qquad(3.9)$$

Proof. If we denote the index of summation in each of the above ${}_4F_3$'s by s and if we shift s to $s-1$ in the second ${}_4F_3$, then we find that the expression L on the left of (3.9) is

$$L=\sum_{s\geqq 0}\frac{(i-2j+1)_s(1-j)_s\left(-j-\frac{\mu}{2}+\frac{1}{2}\right)_s(i+\mu+2j)_s}{s!(i-j+1)_s\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_s s!}$$

$$-(2i+\mu+1)\sum_{s\geqq 0}\frac{s\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_s(i+2j+\mu+1)_{s-1}(-j)_{s+1}(i-2j+2)_{s-1}}{s!(i-j+1)_{s+1}\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_s s!}$$

$$+\left(\frac{\mu}{2}+i+\frac{1}{2}\right)(i+1)\sum_{s\geqq 0}\frac{(i-2j+2)_{s+1}(-j)_{s+1}\left(-j-\frac{\mu}{2}+\frac{3}{2}\right)_s(i+\mu+2j+1)_s}{s!(i-j+1)_{s+2}\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{s+1}(s+1)!}.$$

Since $(2i+\mu+1)=(i+\mu+2j)+(i-2j+1)$, we combine the first sum with the $(i+\mu+2j)$ portion of the second sum:

$$\sum_{s\geqq 0}\frac{(i-2j+2)_{s-1}(1-j)_s\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_s(i+\mu+2j)_s}{s!(i-j+1)_{s+1}\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_s s!}$$
$$\times\{(i-2j+1)(i-j+s+1)-(-j)s\},$$

and since the expression inside $\{\ \}$ is just $(s+i-2j+1)(i-j+1)$ we see that this last expression equals

$$\sum_{s\geqq 0}\frac{(i-2j+2)_s(1-j)_s\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_s(i+\mu+2j)_s}{s!(i-j+2)_s\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_s s!}.$$

Therefore the expression on the left of (3.9) may be written as follows:

$$L=\sum_{s\geqq 0}\frac{(i-2j+2)_s(1-j)_s\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_s(i+\mu+2j)_s}{s!(i-j+2)_s\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_s s!}$$

$$-\sum_{s\geqq 0}\frac{s\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_s(i+2j+\mu+1)_{s-1}(-j)_{s+1}(i-2j+1)_s}{s!\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_s(i-j+1)_{s+1}s!}$$

$$+\tfrac{1}{2}(i+1)(\mu+2i+1)$$

$$\times \sum_{s \geqq 0} \frac{(i-2j+2)_{s+1}(-j)_{s+1}\left(-j-\frac{\mu}{2}+\frac{3}{2}\right)_s (i+\mu+2j+1)_s}{s!\,(i-j+1)_{s+2}\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{s+1}(s+1)!}.$$

Let us now shift the index of summation in the second sum from s to $s+1$ and then combine it with the third sum. Hence

$$L = \sum_{s \geqq 0} \frac{(i-2j+2)_s(1-j)_s\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_s (i+\mu+2j)_s}{s!\,(i-j+2)_s\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_s s!}$$
$$+ \sum_{s \geqq 0} \frac{\left(-j-\frac{\mu}{2}+\frac{3}{2}\right)_s (i+\mu+2j+1)_s(-j)_{s+1}(i-2j+2)_s}{s!\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{s+1}(i-j+1)_{s+2}(s+1)!}$$
$$\times \left[-\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)(-j+s+1)(i-2j+1)+(i+1)\right.$$
$$\left.\times \left(\frac{\mu}{2}+i+\frac{1}{2}\right)(i-2j+s+2)\right].$$

The expression inside "[ ]" factors into

$$(i+1-j)\left\{(1+s-j)(\mu+2j+i)+(i+1)\left(\frac{\mu}{2}+i+\frac{1}{2}\right)\right\}.$$

Hence we conclude that L (the expression on the left side of (3.9)) is given by

$$L = \sum_{s \geqq 0} \frac{(i-2j+2)_s(1-j)_s\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_s (i+\mu+2j)_s}{s!\,(i-j+2)_s\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_s s!}$$
$$+ \sum_{s \geqq 0} \frac{(i-2j+2)_s(-j)_{s+1}\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_s (i+\mu+2j+1)_s}{s!\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{s+1}(i-j+2)_{s+1}(s+1)!}$$
$$\times \left\{(1+s-j)(\mu+2j+i)+(i+1)\left(\frac{\mu}{2}+i+\frac{1}{2}\right)\right\}. \tag{3.10}$$

Equation (3.10) now provides the starting point for the final assault. We propose to prove that if for $0 \leqq n \leqq 2j-2-i$

$$L_n=\sum_{s=2j-2-i-n}^{2j-i-2}\frac{(i-2j+2)_s(1-j)_s\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_s(i+\mu+2j)_s}{s!\,(i-j+2)_s\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_s s!}$$

$$+\sum_{s=2j-2-i-n}^{2j-i-2}\frac{(i-2j+2)_s(-j)_{s+1}\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_s(i+\mu+2j+1)_s}{s!\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{s+1}(i-j+2)_{s+1}(s+1)!}$$

$$\times\left\{(1+s-j)(\mu+2j+i)+(i+1)\left(\frac{\mu}{2}+i+\frac{1}{2}\right)\right\},$$

then

$$L_n=\frac{-(-1)^i(i+\mu+2j+1)_{2j-i-3-n}(i-2j+2)_{n+1}}{n!\,(2j-i-2-n)!\,(i-j+2)_n}$$
$$\times\frac{(1-j)_n\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_n\left(\frac{\mu}{2}+3j-n-\frac{3}{2}\right)}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{n+1}}. \tag{3.11}$$

If we can prove (3.11) for $0\leqq n\leqq 2j-i-2$ this will imply the main result since $L=L_{2j-i-2}$, and by (3.11) we see that $L_{2j-i-2}=0$.

We proceed to treat (3.11) by induction on n. For $n=0$,

$$L_0=\frac{(i-2j+2)_{2j-i-2}(1-j)_{2j-i-2}\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_{2j-i-2}(i+\mu+2j)_{2j-i-2}}{(2j-i-2)!\,(i-j+2)_{2j-i-2}\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{2j-i-2}(2j-i-2)!}$$

$$+\frac{(i-2j+2)_{2j-i-2}(-j)_{2j-i-1}\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_{2j-i-2}(i+\mu+2j+1)_{2j-i-2}}{(2j-i-2)!\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{2j-i-1}(i-j+2)_{2j-i-1}(2j-i-1)!}$$

$$\times\left\{(j-i-1)(\mu+2j+i)+(i+1)\left(\frac{\mu}{2}+i+\frac{1}{2}\right)\right\}$$

$$=\frac{(-1)^i\left(\frac{\mu}{2}+j-\frac{1}{2}\right)(i+\mu+2j)_{2j-i-2}}{(2j-i-2)!\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)}$$

$$-\frac{(-1)^i(i+\mu+2j+1)_{2j-i-2}}{(2j-i-1)!\left(\frac{\mu}{2}+j-\frac{1}{2}\right)}\left\{(\mu+2j-1)\left(j-\frac{i}{2}-\frac{1}{2}\right)\right\}$$

$$=\frac{-(-1)^i(i+\mu+2j+1)_{2j-i-3}}{(2j-i-2)!\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)}$$

$$\times\left[-(i+\mu+2j)\left(\frac{\mu}{2}+j-\frac{1}{2}\right)+(\mu+4j-2)\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)\right]$$

$$=\frac{-(-1)^i(i+\mu+2j+1)_{2j-i-3}(i-2j+2)\left(\frac{\mu}{2}+3j-\frac{3}{2}\right)}{(2j-i-2)!\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)},$$

which is (3.11) in the case $n=0$.

We now assume that (3.11) is proved for all nonnegative integers up to but not including a specific n in the interval $(0, 2j-i-2]$. Hence by the induction hypothesis

$$L_n=\frac{(i-2j+2)_{2j-2-i-n}(1-j)_{2j-2-i-n}}{(2j-2-i-n)!\,(i-j+2)_{2j-2-i-n}}$$

$$\times\frac{\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_{2j-2-i-n}(1+\mu+2j)_{2j-2-i-n}}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{2j-2-i-n}(2j-2-i-n)!}$$

$$+\frac{(i-2j+2)_{2j-2-i-n}(-j)_{2j-1-i-n}}{(2j-2-i-n)!\,(i-j+2)_{2j-1-i-n}}$$

$$\times\frac{\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_{2j-2-i-n}(i+\mu+2j+1)_{2j-2-i-n}}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{2j-1-i-n}(2j-i-n-1)!}$$

$$\times\left\{(j-i-n-1)(\mu+2j+i)+(i+1)\left(\frac{\mu}{2}+i+\frac{1}{2}\right)\right\}$$

$$-\frac{(-1)^i(i+\mu+2j+1)_{2j-i-2-n}(i-2j+2)_n}{(n-1)!\,(2j-i-1-n)!\,(i-j+2)_{n-1}}$$

$$\times\frac{(1-j)_{n-1}\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_{n-1}\left(\frac{\mu}{2}+3j-n-\frac{1}{2}\right)}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_n}$$

$$=\frac{-(-1)^i(i+\mu+2j)_{2j-2-i-n}(1-j)_n\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_{n+1}(i-2j+2)_n}{n!\,(2j-2-i-n)!\,(i-j+2)_n\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{n+1}}$$

$$+\frac{(-1)^i(i+\mu+2j+1)_{2j-2-i-n}(-j)_n\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_{n-1}(i-2j+2)_n}{n!\,(2j-i-n-1)!\,(i-j+2)_n\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_n}$$

$$\times\left\{(j-i-n-1)(\mu+2j+i)+(i+1)\left(\frac{\mu}{2}+i+\frac{1}{2}\right)\right\}$$

$$-\frac{(-1)^i(i+\mu+2j+1)_{2j-i-2-n}(i-2j+2)_n}{(n-1)!\,(2j-1-i-n)(i-j+2)_{n-1}}$$

$$\times\frac{(1-j)_{n-1}\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_{n-1}\left(\frac{\mu}{2}+3j-n-\frac{1}{2}\right)}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_n}$$

$$=\frac{-(-1)^i(i+\mu+2j)_{2j-2-i-n}(1-j)_n\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_{n+1}(i-2j+2)_n}{n!\,(2j-2-i-n)!\,(i-j+2)_n\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{n+1}}$$

$$-\frac{(-1)^i(i+\mu+2j+1)_{2j-2-i-n}(-j)_n\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_{n-1}(i-2j+2)_{n-1}}{n!\,(2j-i-n-2)!\,(i-j+2)_n\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_n}$$

$$\times\left\{(j-i-n-1)(\mu+2j+i)+(i+1)\left(\frac{\mu}{2}+i+\frac{1}{2}\right)\right\}$$

$$+\frac{(-1)^i(i+\mu+2j+1)_{2j-i-2-n}(i-2j+2)_{n-1}}{(n-1)!\,(2j-i-n-2)!\,(i-j+2)_{n-1}}$$

$$\times\frac{(1-j)_{n-1}\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_{n-1}\left(\frac{\mu}{2}+3j-n-\frac{1}{2}\right)}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_n}$$

$$=\frac{-(-1)^i(i+\mu+2j+1)_{2j-i-3-n}}{n!\,(2j-i-n-2)!\,(i-j+2)_n}$$

$$\times\frac{(i-2j+2)_{n-1}(1-j)_{n-1}\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_{n-1}}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{n+1}}$$

$$\times\left[(i+\mu+2j)(n-j)\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)(i-2j+n+1)\left(-\frac{\mu}{2}-j+n+\frac{1}{2}\right)\right.$$

$$+(\mu+4j-n-2)(-j)\left(i-j+\frac{\mu}{2}+\frac{3}{2}+n\right)$$

$$\times\left\{(j-i-n-1)(\mu+2j+i)+(i+1)\left(\frac{\mu}{2}+i+\frac{1}{2}\right)\right\}$$

$$\left.-(\mu+4j-n-2)\left(\frac{\mu}{2}+3j-n-\frac{1}{2}\right)n(i-j+n+1)\left(i-j+\frac{\mu}{2}+\frac{3}{2}+n\right)\right].$$

Now we must simplify the expression inside the square brackets. When we multiply it out, we find that there are 752 separate terms initially. After cancellation we find that

$$\begin{aligned}
L_n &= \frac{(-1)^i(i+\mu+2j+1)_{2j-i-3-n}(i-2j+2)_{n-1}(1-j)_{n-1}\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_{n-1}}{n!\,(2j-i-n-2)!\,(i-j+2)_n\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{n+1}} \\
&\quad\times[\tfrac{3}{2}j-\tfrac{3}{2}n+\tfrac{9}{4}ji+\tfrac{39}{2}jin-15jin\mu+\tfrac{3}{2}jin\mu^2-3ji\mu \\
&\quad+33jin^2-10jin^2\mu+14jin^3+\tfrac{3}{4}ji\mu^2+\tfrac{67}{4}jn-15jn\mu \\
&\quad+\tfrac{9}{4}jn\mu^2-2j\mu+\tfrac{3}{4}ji^2+5ji^2n-3ji^2n\mu-ji^2\mu+5ji^2n^2 \\
&\quad+\tfrac{1}{4}ji^2\mu^2+\tfrac{163}{4}jn^2-20jn^2\mu+\tfrac{5}{4}jn^2\mu^2+34jn^3-7jn^3\mu \\
&\quad+9jn^4+\tfrac{1}{2}j\mu^2-\tfrac{9}{4}in+3in\mu-\tfrac{3}{4}in\mu^2-\tfrac{15}{2}in^2+5in^2\mu \\
&\quad-\tfrac{1}{2}in^2\mu^2-7in^3+2in^3\mu-2in^4+2n\mu-\tfrac{1}{2}n\mu^2-12j^2i-47j^2in \\
&\quad+16j^2in\mu+10j^2i\mu-34j^2in^2-j^2i\mu^2-59j^2n+32j^2n\mu-2j^2n\mu^2 \\
&\quad+10j^2\mu-3j^2i^2-7j^2i^2n+2j^2i^2\mu-82j^2n^2+18j^2n^2\mu-31j^2n^3-\tfrac{3}{2}j^2\mu^2 \\
&\quad+21j^3i+34j^3in-8j^3i\mu+83j^3n-20j^3n\mu-16j^3\mu+3j^3i^2+51j^3n^2 \\
&\quad+j^3\mu^2-12j^4i-40j^4n+8j^4\mu-\tfrac{3}{4}i^2n+i^2n\mu-\tfrac{1}{4}i^2n\mu^2-2i^2n^2 \\
&\quad+i^2n^2\mu-i^2n^3+5n^2\mu-\tfrac{3}{4}n^2\mu^2+4n^3\mu-\tfrac{1}{4}n^3\mu^2+n^4\mu-\tfrac{21}{2}j^2 \\
&\quad+27j^3-30j^4+12j^5-\tfrac{25}{4}n^2-\tfrac{35}{4}n^3-5n^4-n^5] \\
&= \frac{(-1)^i(i+\mu+2j+1)_{2j-i-3-n}(i-2j+2)_{n-1}(1-j)_{n-1}\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_{n-1}}{n!\,(2j-i-n-2)!\,(i-j+2)_n\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{n+1}} \\
&\quad\times\left[(i-2j+n+1)(i-2j+n+2)(n-j)\left(-\frac{\mu}{2}-j+n+\frac{1}{2}\right)\left(\frac{\mu}{2}+3j-n-\frac{3}{2}\right)\right] \\
&= \frac{-(-1)^i(i+\mu+2j+1)_{2j-i-3-n}(i-2j+2)_{n+1}}{n!\,(2j-i-n-2)!\,(i-j+2)_n} \\
&\quad\times\frac{(1-j)_n\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_n\left(\frac{\mu}{2}+3j-n-\frac{3}{2}\right)}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{n+1}},
\end{aligned} \tag{3.12}$$

and so (3.11) is proved by mathematical induction. As we remarked earlier (3.11) implies (3.9) the assertion of our theorem. □

4. Proof of the Weak Conjectures

We begin with a matrix identity that will immediately imply Theorem 8.

Theorem 7. *The following matrix identity holds* (i *and* j *lie in* $[0, m-1]$)

$$\left(\delta_{ij}+\binom{\mu+i+j}{i}\right)\times(e_{ij})=(f_{ij}), \tag{4.1}$$

where

$$e_{ij}=0 \quad \textit{for} \quad i>j \tag{4.2}$$

$$e_{ii}=1, \tag{4.3}$$

and more generally for $2j \geqq i$

$$e_{i,\,2j}=\sum_{s\geqq 0}(-1)^i\binom{j-s}{2j-2s-i}\frac{(j-s)_{2s}(-\mu-3j-1)_s 4^{-s}}{s!\left(-\frac{\mu}{2}-j\right)_s\left(-\frac{\mu}{2}-2j+\frac{1}{2}\right)_s}, \tag{4.4}$$

and for $2j-1 \geqq i$

$$e_{i,\,2j-1}=\frac{(-1)^{i-1}}{2}\sum_{s\geqq 0}\left(\binom{j-s}{2j-1-i-2s}+\binom{j-s-1}{2j-1-i-2s}\right)\frac{(j-s)_{2s}(-\mu-3j+1)_s 4^{-s}}{s!\left(-\frac{\mu}{2}-j+1\right)_s\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_s}\times\frac{\mu+3j-1-3s}{\mu+3j-1}; \tag{4.5}$$

furthermore

$$f_{ij}=0 \quad \textit{for} \quad j>i, \tag{4.6}$$

and

$$f_{ii}=\Delta_i(\mu), \tag{4.7}$$

where $\Delta_i(\mu)$ *is defined in the statement of Theorem* 8 *in Section* 1.

Proof. Since there is no assertion about the f_{ij} for $j<i$, we need only show that the assumption of (4.2), (4.3), (4.4) and (4.5) implies (4.6) and (4.7). We note that (4.2) and (4.3) are implicit in (4.4) and (4.5).

We begin by treating $f_{i,\,2j}$:

$$\begin{aligned}
f_{i,\,2j}&=\sum_{l\geqq 0}\left\{\delta_i+\binom{\mu+i+l}{i}\right\}e_{l,\,2j}\\
&=\sum_{l\geqq 0}\left\{\delta_i+\binom{\mu+i+l}{i}\right\}\\
&\quad\times\sum_{s\geqq 0}(-1)^l\binom{j-s}{2j-2s-l}\frac{(j-s)_{2s}(-\mu-3j-1)_s 4^{-s}}{s!\left(-\frac{\mu}{2}-j\right)_s\left(-\frac{\mu}{2}-2j+\frac{1}{2}\right)_s}\\
&=\sum_{s\geqq 0}\frac{(j-s)_{2s}(-\mu-3j-1)_s 4^{-s}}{s!\left(-\frac{\mu}{2}-j\right)_s\left(-\frac{\mu}{2}-2j+\frac{1}{2}\right)_s}\sum_{l\geqq 0}(-1)^l\left\{\binom{\mu+i+l}{i}+\delta_{il}\right\}\binom{j-s}{2j-2s-l}\\
&=\sum_{s\geqq 0}\frac{(j-s)_{2s}(-\mu-3j-1)_s 4^{-s}}{s!\left(-\frac{\mu}{2}-j\right)_s\left(-\frac{\mu}{2}-2j+\frac{1}{2}\right)_s}\left\{\binom{\mu+i+j-s}{i-j+s}+(-1)^i\binom{j-s}{2j-2s-i}\right\}
\end{aligned}$$

(by the Chu-Vandermonde summation [7; p. 3])

$$
\begin{aligned}
&=(M(i+1,j,\mu+1;0,1)-M(i,j,\mu+1;0,1))\\
&\quad+(\Lambda(i,j,\mu+1;0,1)-\Lambda(i+1,j,\mu+1;0,1))\\
&=(M(i+1,j,\mu+1;0,1)-\Lambda(i+1,j,\mu+1;0,1))\\
&\quad-(M(i,j,\mu+1;0,1)-\Lambda(i,j,\mu+1;0,1))\\
&=\begin{cases}0 & \text{for } 0\leqq i\leqq 2j-1 \text{ by Theorem 5}\\ M(2j+1,j,\mu+1;0,1) & \text{for } i=2j \text{ by Theorem 5 and Corollary 5a}\end{cases}\\
&=\begin{cases}0 \quad \text{for } 0\leqq i\leqq 2j-1\\ 2 \quad \text{for } i=j=0\\ \dfrac{(\mu+2j+2)_j\left(\frac{\mu}{2}+2j+\frac{3}{2}\right)_{j-1}}{(j)_j\left(\frac{\mu}{2}+j+\frac{3}{2}\right)_{j-1}} & \text{if } i=2j>0\end{cases}\\
&=\begin{cases}0 & \text{for } 0\leqq i\leqq 2j-1\\ \Delta_{2j}(\mu) & \text{for } i=2j,\end{cases}
\end{aligned}
\tag{4.8}
$$

where $\Delta_{2j}(\mu)$ is defined by (1.14).

Next we treat $f_{i,\,2j-1}$:

$$
\begin{aligned}
f_{i,\,2j-1}&=\sum_{l\geqq 0}\left\{\delta_{il}+\binom{\mu+i+l}{i}\right\}e_{l,\,2j-1}\\
&=\tfrac{1}{2}\sum_{l\geqq 0}\left\{\binom{m+i+l}{i}+\delta_{il}\right\}\sum_{s\geqq 0}(-1)^{l-1}\left(\binom{j-s}{2j-1-l-2s}+\binom{j-s-1}{2j-1-l-2s}\right)\\
&\quad\times\frac{(j-s)_{2s}(-\mu-3j+1)_s 4^{-s}}{s!\left(-\frac{\mu}{2}-j+1\right)_s\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_s}\;\frac{\mu+3j-1-3s}{\mu+3j-1}\\
&=-\tfrac{1}{2}\sum_{s\geqq 0}\frac{(j-s)_{2s}(-\mu-3j+1)_s 4^{-s}}{s!\left(-\frac{\mu}{2}-j+1\right)_s\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_s}\;\frac{\mu+3j-1-3s}{\mu+3j-1}\\
&\quad\times\sum_{l\geqq 0}(-1)^l\left\{\binom{m+i+l}{i}+\delta_{il}\right\}\left\{\binom{j-s}{2j-1-l-2s}+\binom{j-s-1}{2j-1-l-2s}\right\}\\
&=-\tfrac{1}{2}\sum_{s\geqq 0}\left\{(-1)^i\binom{j-s}{2j-2s-i-1}+(-1)^i\binom{j-s-1}{2j-2s-i-1}\right.\\
&\quad\left.-\binom{\mu+i+j-s-1}{i-j+s}-\binom{\mu+i+j-s}{i-j+s+1}\right\}\\
&\quad\times\frac{(j-s)_{2s}(-\mu-3j+1)_s 4^{-s}(-\mu-3j+1+3s)}{s!\left(-\frac{\mu}{2}-j+1\right)_s\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_s(-\mu-3j+1)}\\
&=-\tfrac{1}{2}\sum_{s\geqq 0}\left\{(-1)^i\binom{j-s}{2j-2s-i-1}+(-1)^i\binom{j-s-1}{2j-2s-i-1}\right.\\
&\quad\left.-\binom{\mu+i+j-s-1}{i-j+s}-\binom{\mu+i+j-s}{i-j+s+1}\right\}
\end{aligned}
$$

$$\times\frac{(j-s)_{2s}(-\mu-3j+1)_s 4^{-s}}{s!\left(-\frac{\mu}{2}-j+1\right)_s\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_s}$$

$$-\tfrac{3}{2}\sum_{s\geqq 0}\left\{(-1)^i\binom{j-s-1}{2j-2s-i-3}+(-1)^i\binom{j-s-2}{2j-2s-i-3}\right.$$

$$\left.-\binom{\mu+i+j-s-2}{i-j+s+1}-\binom{\mu+i+j-s-1}{i-j+s+2}\right\}$$

$$\times\frac{(j-s-1)_{2s+2}(-\mu-3j+2)_s 4^{-s-1}}{s!\left(-\frac{\mu}{2}-j+1\right)_{s+1}\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_{s+1}} \tag{4.9}$$

(here the single sum has been split into two by $(-\mu-3j+1+3s)=(-\mu-3j+1)+(3s)$)

$$=-\tfrac{1}{2}\sum_{s\geqq 0}\left\{(-1)^i\binom{j-s}{2j-2s-i-1}+(-1)^i\binom{j-s-1}{2j-2s-i-1}\right.$$

$$\left.-\binom{\mu+i+j-s-1}{i-j+s}-\binom{\mu+i+j-s}{i-j+s+1}\right\}$$

$$\times\frac{(j-s)_{2s}(-\mu-3j+1)_s 4^{-s}}{s!\left(-\frac{\mu}{2}-j+1\right)_s\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_s}$$

$$-\tfrac{3}{2}\sum_{s\geqq 0}\left\{(-1)^i\binom{j-s-1}{2j-2s-i-3}+(-1)^i\binom{j-s-2}{2j-2s-i-3}\right\}$$

$$\times\frac{(j-s-1)_{2s+2}(-\mu-3j+2)_s 4^{-s-1}}{s!\left(-\frac{\mu}{2}-j+1\right)_{s+1}\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_{s+1}}$$

$$+\tfrac{3}{2}\sum_{s\geqq 0}\left\{\binom{\mu+i+j-s-2}{i-j+s+1}+\binom{\mu+i+j-s-1}{i-j+s+2}\right\}$$

$$\times\frac{(j-s-1)_{2s+2}(-\mu-3j+2)_s 4^{-s-1}}{s!\left(-\frac{\mu}{2}-j+1\right)_{s+1}\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_{s+1}}$$

$$=-S_1-S_2+S_3 \text{ say.}$$

Now

$$\begin{aligned}S_1 &= \tfrac{1}{2}(-(\Lambda(i+1,j,\mu-1;0,1)-\Lambda(i+2,j,\mu-1;0,1))\\ &\quad -\Lambda(i+1,j,\mu-1;0,1)-M(i+1,j,\mu-1;0,1)\\ &\quad -M(i+2,j,\mu-1;0,1))\\ &=\begin{cases}-\tfrac{3}{2}\Lambda(i+1,j,\mu-1;0,1) & \text{for } 0\leqq i\leqq 2j-2\\ -\tfrac{3}{2}\Lambda(i+1,j,\mu-1;0,1)-\tfrac{1}{2}M(2j+1,j,\mu-1;0,1) & \text{for } i=2j-1\end{cases}\\ &=\begin{cases}-\tfrac{3}{2}\Lambda(i+1,j,\mu-1;0,1) & \text{for } 0\leqq i\leqq 2j-2\\ -\tfrac{3}{2}-\tfrac{1}{2}M(2j+1,j,\mu-1;0,1) & \text{for } i=2j-1 \text{ (by Corollary 5a).}\end{cases}\end{aligned}$$

As for S_2, we see that

$$S_2 = +\tfrac{3}{2}(i+1)(-1)^i \times \sum_{s\geqq 0} \binom{j-s-1}{2j-2s-i-3} \frac{(j-s)_{2s+1}(-\mu-3j+2)_s\, 4^{-s-1}}{s!\left(-\frac{\mu}{2}-j+1\right)_{s+1}\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_{s+1}}$$

$$= -\tfrac{3}{2}(i+1)\,\frac{(2j-i-2)\,A(i,j,\mu-3;\,3,\,-1)}{4\left(-\frac{\mu}{2}-j+1\right)\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)}.$$

Finally we treat S_3:

$$S_3 = \tfrac{3}{2}(\mu+2i+1)\sum_{s\geqq 0}\frac{(\mu+i+j-s-2)!\,(j-s-1)_{2s+2}}{(i-j+s+2)!\,(\mu+2j-2s-3)!\,s!} \times \frac{(-\mu-3j+2)_s\,4^{-s-1}}{\left(-\frac{\mu}{2}-j+1\right)_{s+1}\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_{s+1}}$$

$$= \frac{3(\mu+2i+1)\,2^{-1}(\mu+i+j-2)!\,j(-j+1)}{2\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)(i-j+2)!\,(\mu+2j-2)!} \times \sum_{s\geqq 0}\frac{(-j+2)_s(j+1)_s\left(-\frac{\mu}{2}-j+\frac{3}{2}\right)_s(-\mu-3j+2)_s}{s!\left(-\frac{\mu}{2}-2j+\frac{5}{2}\right)_s(i-j+3)_s(-\mu-i-j+2)_s}$$

$$= \frac{3(\mu+2i+1)(\mu+i+j-2)!\,j(j-1)}{2(\mu+4j-3)(i-j+2)!\,(\mu+2j-2)!} \times {}_4F_3\left[\begin{matrix}-j+2,\, j+1,\, -\frac{\mu}{2}-j+\frac{3}{2},\, -\mu-3j+2;\ 1\\ -\frac{\mu}{2}-2j+\frac{5}{2},\, i-j+3,\, -\mu-i-j+2\end{matrix}\right]$$

$$= \frac{3(\mu+2i+1)(\mu+i+j-2)!\,j(j-1)}{2(\mu+4j-3)(i-j+2)!\,(\mu+2j-2)!}\,\frac{\left(i-j+\frac{\mu}{2}+\frac{5}{2}\right)_{j-2}(-j+1)_{j-2}}{(\mu+i+1)_{j-2}\left(-\frac{\mu}{2}-2j+\frac{5}{2}\right)_{j-2}} \times {}_4F_3\left[\begin{matrix}-\frac{\mu}{2}-j+\frac{3}{2},\, i+2j+\mu+1,\, -j+2,\, i-2j+2;\ 1\\ i-j+\frac{\mu}{2}+\frac{5}{2},\, i-j+3,\, 2\end{matrix}\right].$$

Now when $i=2j-1$ we see that $S_2=0$. Consequently

$$f_{2j-1,\,2j-1}=\tfrac{3}{2}+\tfrac{1}{2}M(2j+1,j,\mu-1;\,0,1)$$
$$+\tfrac{3}{2}\sum_{s\geqq 0}\frac{\left(-\frac{\mu}{2}-j+\frac{1}{2}\right)_{s+1}(\mu+4j-1)_{s+1}(-j+1)_{s+1}}{(s+1)!\left(j+\frac{\mu}{2}+\frac{1}{2}\right)_{s+1}(j+1)_{s+1}}$$
$$=\tfrac{1}{2}M(2j+1,j,\mu-1;\,0,1)$$
$$+\tfrac{3}{2}\,{}_3F_2\left[\begin{matrix}-j+1,\ \mu+4j-1,\ -\frac{\mu}{2}-j+\frac{1}{2};\ 1\\ j+\frac{\mu}{2}+\frac{1}{2},\ j+1\end{matrix}\right]$$
$$=\frac{1}{2}\,\frac{(\mu+2j)_j\left(2j+\frac{\mu}{2}+\frac{1}{2}\right)_{j-1}}{(j)_j\left(\frac{\mu}{2}+j+\frac{1}{2}\right)_{j-1}}+\frac{3}{2}\,\frac{(\mu+2j)_{j-1}\left(-\frac{\mu}{2}-3j+\frac{3}{2}\right)_{j-1}}{\left(j+\frac{\mu}{2}+\frac{1}{2}\right)_{j-1}(-2j+1)_{j-1}}$$
$$=\frac{1}{2}\,\frac{(\mu+2j)_{j-1}\left(\frac{\mu}{2}+2j+\frac{1}{2}\right)_{j-1}}{(j)_j\left(\frac{\mu}{2}+j+\frac{1}{2}\right)_{j-1}}(\mu+3j-1+3j)$$
$$=\frac{(\mu+2j)_{j-1}\left(\frac{\mu}{2}+2j+\frac{1}{2}\right)_j}{(j)_j\left(\frac{\mu}{2}+j+\frac{1}{2}\right)_{j-1}}=\Delta_{2j-1}(\mu)$$

by (1.15). Thus we have now established (4.7) for all $i\geqq 0$.

Finally for $i<2j-1$, we see that

$$f_{i,\,2j-1}=-S_1-S_2+S_3$$
$$=\tfrac{3}{2}\Delta(i+1,j,\mu-1;\,0,1)$$
$$+\frac{3}{2}\,\frac{(i+1)(2j-i-2)\,\Delta(i,j,\mu-3;\,3,-1)}{4\left(-\frac{\mu}{2}-j+1\right)\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)}$$
$$+\frac{3(\mu+2i+1)(\mu+i+j-2)!\,j(j-1)\left(i-j+\frac{\mu}{2}+\frac{5}{2}\right)_{j-2}(1-j)_{j-2}}{(\mu+4j-3)(i-j+2)!\,(\mu+2j-2)!\,(\mu+i+1)_{j-2}\left(-\frac{\mu}{2}-2j+\frac{5}{2}\right)_{j-2}}$$
$$\times\,{}_4F_3\left[\begin{matrix}-\frac{\mu}{2}-j+\frac{3}{2},\ i+2j+\mu+1,\ -j+2,\ i-2j+2;\ 1\\ i-j+\frac{\mu}{2}+\frac{5}{2},\ i-j+3,\ 2\end{matrix}\right].$$

For the two Λ-functions we use the second ${}_4F_3$ expression that arises in the string of equations equating $M(i,j,m;a,\omega)$ with $\Lambda(i,j,m;a,\omega)$ in the proof of Theorem 5. Hence

$$f_{i,\,2j-1}=\frac{3(\mu+i)!\,(1-j)_{j-1}\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)_{j-1}}{2(i-j)!\,(\mu+2j-1)!\left(-\frac{\mu}{2}-2j+\frac{3}{2}\right)_{j-1}}$$

$$\times\left\{{}_4F_3\left[\begin{matrix}i-2j+1,\,1-j,\,-j-\frac{\mu}{2}+\frac{1}{2},\,i+\mu+2j;\,1\\ i-j+1,\,i-j+\frac{\mu}{2}+\frac{3}{2},\,1\end{matrix}\right]\right.$$

$$+\frac{(2i+\mu+1)\left(\frac{\mu}{2}+j-\frac{1}{2}\right)j(j-1)}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)(i-j+1)(i-j+2)}$$

$$\times\,{}_4F_3\left[\begin{matrix}i-2j+2,\,2-j,\,-j-\frac{\mu}{2}+\frac{3}{2},\,i+\mu+2j+1;\,1\\ i-j+3,\,i-j+\frac{\mu}{2}+\frac{5}{2},\,2\end{matrix}\right]$$

$$+\frac{\left(\frac{\mu}{2}+i+\frac{1}{2}\right)(i+1)(2j-i-2)j}{\left(i-j+\frac{\mu}{2}+\frac{3}{2}\right)(i-j+1)(i-j+2)}$$

$$\left.\times\,{}_4F_3\left[\begin{matrix}i-2j+3,\,1-j,\,-j-\frac{\mu}{2}+\frac{3}{2},\,i+\mu+2j+1;\,1\\ i-j+3,\,i-j+\frac{\mu}{2}+\frac{5}{2},\,2\end{matrix}\right]\right\}$$

$$=0,\qquad \text{by Theorem 6.}$$

It appears at first glance that we can invoke Theorem 6 for i and j integers only for $i\geqq j$. However the above expression is indentically zero provided $i-2j+2$ is a nonpositive integer and $i-j$ is any complex (or real) number other than a negative integer. The conclusion for all integral i with $0\leqq i<2j-1$ then follows by continuity.

As a direct corollary of Theorem 7, we immediately deduce:

Theorem 8.

$$\det\left(\delta_{ij}+\binom{\mu+i+j}{i}\right)_{0\leqq i,\,j\leqq m-1}=\prod_{j=0}^{m-1}\Delta_j(\mu),$$

where

$$\Delta_{2j}(\mu)=\begin{cases}\dfrac{(\mu+2j+2)_j\left(\dfrac{\mu}{2}+2j+\dfrac{3}{2}\right)_{j-1}}{(j)_j\left(\dfrac{\mu}{2}+j+\dfrac{3}{2}\right)_{j-1}}, & j>0\\ 2, \quad j=0,\end{cases} \tag{4.10}$$

and

$$\Delta_{2j-1}(\mu)=\frac{(\mu+2j)_{j-1}\left(\dfrac{\mu}{2}+2j+\dfrac{1}{2}\right)_j}{(j)_j\left(\dfrac{\mu}{2}+j+\dfrac{1}{2}\right)_{j-1}}, \quad j>0 \tag{4.11}$$

Proof. By Theorem 7, we see that

$$\begin{aligned}\det\left(\delta_{ij}+\binom{\mu+i+j}{i}\right)_{0\leq i,j\leq m-1}&=\frac{\det(f_{ij})}{\det(e_{ij})}\\&=f_{00}f_{11}\cdots f_{m-1,m-1}\\&=\prod_{j=0}^{m-1}\Delta_j(\mu). \quad \square\end{aligned}$$

As a consequence of Theorem 8, we can now prove the weak forms of both conjectures given in the introduction.

Theorem 9 (The Weak Macdonald Conjecture). *The total number of plane partitions π such that $D(\pi)\subseteq\mathscr{B}_{m,m,m}$ and $D(\pi)$ is G_3 invariant is*

$$\prod_{i=1}^{m}\left\{\frac{3i-1}{3i-2}\prod_{j=i}^{m}\frac{m+i+j-1}{2i+j-1}\right\}.$$

Proof. We let γ_m denote the product given in Theorem 9. Then by Theorem 4 with $q=1$ and Theorem 8, we see that we must prove that

$$\gamma_1=\Delta_0(0), \tag{4.12}$$

$$\frac{\gamma_m}{\gamma_{m-1}}=\Delta_{m-1}(0). \tag{4.13}$$

Now (4.12) is clear since $\gamma_1=\frac{2}{1}\cdot\frac{2}{2}=2=\Delta_0(0)$. As for (4.13), we see that

$$\begin{aligned}\frac{\gamma_m}{\gamma_{m-1}}&=\frac{3m-1}{3m-2}\times\prod_{i=1}^{m}\frac{2m+i-1}{2i+m-1}\times\prod_{i=1}^{m-1}\frac{2m+i-2}{m+2i-2}=\frac{3m-1}{3m-2}\,\frac{(2m)_m(2m-1)_{m-1}}{(m)_{2m-2}(3m-1)}\\&=\frac{3m-1}{3m-2}\times\frac{(2m)_{m-1}(2m-1)_{m-1}}{(m)_{2m-2}}=\frac{(3m-1)}{(3m-2)}\times\frac{(2m)_{m-1}}{(m)_{m-1}},\end{aligned}$$

and by (1.15)

$$\begin{aligned}\Delta_{2j-1}(0)&=\frac{(2j)_{j-1}(2j+\frac{1}{2})_j}{(j)_j(j+\frac{1}{2})_{j-1}}=\frac{(3j-\frac{1}{2})(4j)_{2j-1}}{(3j-1)(2j)_{2j-1}}\\&=\frac{6j-1}{6j-2}\,\frac{(4j)_{2j-1}}{(2j)_{2j-1}}=\frac{\gamma_{2j}}{\gamma_{2j-1}},\end{aligned}$$

while by (1.14)

$$\begin{aligned}\Delta_{2j}(0)&=\frac{(2j+2)_j(2j+\frac{3}{2})_{j-1}}{(j)_j(j+\frac{3}{2})_{j-1}}=\frac{3j+1}{3j+\frac{1}{2}}\times\frac{(2j+2)_{j-1}(2j+\frac{3}{2})}{(j)_j(j+\frac{3}{2})_{j-1}}\\&=\frac{(3j+1)}{(3j+\frac{1}{2})}\times\frac{(2j+1)_j(2j+\frac{3}{2})_j}{(j+1)_j(j+\frac{1}{2})_j}=\frac{(6j+2)}{(6j+1)}\times\frac{(4j+2)_{2j}}{(2j+1)_{2j}}=\frac{\gamma_{2j+1}}{\gamma_{2j}}.\end{aligned}$$

Thus Theorem 9 is established. □

Theorem 10. *The total number of descending plane partitions whose parts do not exceed m is* $\prod_{i=1}^{m}\prod_{j=1}^{m}\frac{m+i+j-1}{2i+j-1}$.

Proof. We let δ_m denote the product given in Theorem 10. Then by Theorem 3′ with $d=0$, $q=1$ and Theorem 8, we see that we must establish

$$\delta_2=\Delta_0(2), \tag{4.14}$$

$$\frac{\delta_m}{\delta_{m-1}}=\Delta_{m-2}(2). \tag{4.15}$$

Now (4.14) is clear since $\delta_2=\frac{3}{2}\cdot\frac{4}{3}\cdot\frac{5}{5}=2=\Delta_0(2)$. As for (4.15), we see that

$$\frac{\delta_m}{\delta_{m-1}}=\frac{3m-2}{3m-1}\cdot\frac{\gamma_m}{\gamma_{m-1}}=\frac{(2m)_{m-1}}{(m)_{m-1}},$$

and by (1.15)

$$\begin{aligned}\Delta_{2j-1}(2)&=\frac{(2j+2)_{j-1}(2j+\frac{3}{2})_j}{(j)_j(j+\frac{3}{2})_{j-1}}=\frac{(2j+1)_j(2j+\frac{3}{2})_j}{(j+1)_j(j+\frac{1}{2})_j}\\&=\frac{(4j+2)_{2j}}{(2j+1)_{2j}}=\frac{\delta_{2j+1}}{\delta_{2j}},\end{aligned}$$

while by (1.14)

$$\begin{aligned}\Delta_{2j}(2)&=\frac{(2j+4)_j(2j+\frac{5}{2})_{j-1}}{(j)_j(j+\frac{5}{2})_{j-1}}\\&=\frac{(2j+3)_{j+1}(2j+\frac{5}{2})_{j-1}}{(j+1)_j(j+\frac{3}{2})_j}=\frac{(2j+3)_{j+1}(2j+\frac{3}{2})_j}{(j+1)_j(j+\frac{3}{2})_{j+1}}\\&=\frac{(2j+2)_{j+1}(2j+\frac{5}{2})_j}{(j+1)_{j+1}(j+\frac{3}{2})}\times\frac{(2j+1)(3j+3)}{(2j+2)(3j+\frac{3}{2})}\\&=\frac{(4j+2)_{2j+1}}{(2j+2)_{2j+1}}=\frac{\delta_{2j+2}}{\delta_{2j+1}}.\end{aligned}$$

Thus Theorem 10 is established. □

Clearly we can apply Theorem 8 to the case $q=1$ of Theorem 3′; in [5] we discuss this application in more detail.

5. Conclusion

Obviously the final assault on Macdonald's conjecture is the first order of business now. One is tempted to hope that the q-analogs of the results on hypergeometric series utilized in this paper will be adequate to treat both Macdonald's conjecture and the descending plane partitions conjecture. However just as the full q-series proof of MacMahon's conjecture [4] was significantly more complicated than just the case $q=1$ [1], so too is it clear upon inspection that the full conjectures studied here are not obtained purely by direct substitution of q-analogs of Theorems 5 and 6. A discussion of the evidence for these conjectures and related problems occurs in [5].

In [2] the equivalence of the Bender-Knuth conjecture with MacMahon's conjecture was proved using only elementary row operations on determinants. So far similar efforts to prove the equivalence of Macdonald's conjecture with the descending plane partitions conjecture have gone astray.

The assertion in Theorem 5 is the following hypergeometric series identity:

$$\begin{aligned}
&\frac{\left(\frac{m}{2}+2j\right)_{a+\omega-1}}{\left(\frac{m}{2}+j+1\right)_{a+\omega-1}}\binom{m+i+j+a-1}{i-j+a-1}\\
&\quad\times {}_4F_3\left[\begin{matrix}1-j, j, -m-3j-a-\omega+1, -\frac{m}{2}-j; 1\\ i-j+a, -\frac{m}{2}-2j-a-\omega+2, -m-i-j-a+1\end{matrix}\right]\\
&=(-1)^{i+a}\binom{j+a+\omega-2}{i-j+a-1}\\
&\quad\times {}_4F_3\left[\begin{matrix}-j+\frac{i+a}{2}, -j+\frac{i+a+1}{2}, j+a+\omega-1, -m-3j-a-\omega+1; 1\\ i-j+a, -\frac{m}{2}-2j+1, -\frac{m}{2}-j+\frac{1}{2}\end{matrix}\right].
\end{aligned} \tag{5.1}$$

Since the crucial step of the proof of this identity is obtained from applying (3.6) when the series on the left is not only nearly poised of the second kind but also balanced, it would appear that such series have a significant amount of structure beyond that of merely balanced ${}_4F_3$'s. Recent work of J. Wilson [19] and [20] and R. Askey and J. Wilson [6] has made clear the importance of balanced ${}_4F_3$'s.

Theorem 6 is essentially a contiguous relation among three balanced ${}_4F_3$'s. Consequently it should be derivable from the ${}_4F_3$ contiguous relations described by J. Wilson [19]. Indeed, R. Askey has outlined how the derivation should proceed from the work in Chapter 4 of Wilson's thesis [20]. Equation (3.11) essentially asserts that the tails of the three series in Theorem 6 are indeed summable to the right side of (3.11). Thus we may deduce from (3.11) and (3.9) that the sum of three truncated balanced ${}_4F_3$'s of this type is summable. Askey also suggests that Wilson's work can be used to explain this phenomenon.

Finally we mention that the "massive" computation in (3.12) for L_n which was essential to the proof of Theorem 6 was done by the FORMAC Utility Program due to H.D. Noble of the Pennsylvania State University Computation Center.

References

1. Andrews, G.E.: MacMahon's conjecture on symmetric plane partitions. Proc. Nat. Acad. Sci. USA, **74**, 426–429 (1977)
2. Andrews, G.E.: Plane partitions (II): the equivalence of the Bender-Knuth and MacMahon conjectures. Pac. J. Math., **72**, 283–291 (1977)
3. Andrews, G.E.: The Theory of Partitions, Addision-Wesley, Reading, 1976
4. Andrews, G.E.: Plane partitions (I): the MacMahon conjecture, Advances in Math. (to appear)
5. Andrews, G.E.: On Macdonald's conjecture and descending plane partitions, Proceedings of the Alfred Young Day Conference (to appear)
6. Askey, R., Wilson, J.: A set of orthogonal polynomials that generalize the Racah coefficients or 6-j symbols, S.I.A.M. J. Math. Anal., (to appear)
7. Bailey, W.N.: Generalized Hypergeometric Series. London and New York: Cambridge University Press 1935
8. Carlitz, L.: Rectangular arrays and plane partitions, Acta Arith., **13**, 22–47 (1967)
9. Gansner, E.: Matrix correspondences and the enumeration of plane partitions, Ph.D. Thesis, M.I.T., 1978
10. Macdonald, I.G.: Lecture on plane partitions, Oberwolfach Conference on Combinatorics and Special Functions, May 1977
11. MacMahon, P.A.: Partitions of numbers whose graphs possess symmetry. Trans. Cambridge Phil. Soc., **17**, 149–170 (1898–99)
12. MacMahon, P.A.: Combinatory Analysis, Vol. 2. London and New York: Cambridge University Press 1916
13. Muir, T.: A Treatise on the Theory of Determinants. London: Longmans 1933
14. Stanley, R.P.: Theory and applications of plane partitions I, Studies in Appl. Math., **50**, 167–188 (1971)
15. Stanley, R.P.: Theory and applications of plane partitions II, Studies in Appl. Math., **50**, 259–279 (1971)
16. Whipple, F.J.W.: Well-poised series and other generalized hypergeometric series, Proc. London Math. Soc. (2), **25**, 525–544 (1926)
17. Whipple, F.J.W.: Some transformations of generalized hypergeometric series, Proc. London Math. Soc. (2), **26**, 257–272 (1927)
18. Whittaker, E.T., Watson, G.N.: A Course of Modern Analysis, 4th ed., London and New York: Cambridge University Press 1958
19. Wilson, J.: Three-term contiguous relations and some new orthogonal polynomials, from Pade and Rational Approximation (Saff and Varga, eds.). New York-London: Academic Press, 1977
20. Wilson, J.: Ph.D. Thesis, University of Wisconsin, 1978

Received October 3, 1978

Inventiones math. 53, 227–247 (1979)

Inventiones mathematicae

Closures of Conjugacy Classes of Matrices are Normal

Hanspeter Kraft[1] and Claudio Procesi[2]

[1] Sonderforschungsbereich Theoretische Mathematik, Universität Bonn D-5300 Bonn, Federal Republic of Germany
[2] Istituto di Matematica, Universitá di Roma, I-Rome, Italy

0. Introduction

0.1. The purpose of this paper is to prove the following theorem: Let A be an $n \times n$ matrix over an algebraically closed field K of characteristic zero, C_A the conjugacy class of A and $\overline{C_A}$ its (Zariski-) closure.

Theorem. $\overline{C_A}$ *is normal, Cohen-Macaulay with rational singularities.*

If a variety X with a G-action (G reductive) is the closure of an orbit $\mathcal{O}$ and $\dim(X \setminus \mathcal{O}) \leqq \dim X - 2$, it is a crucial question for the geometry of X to decide whether the singularity (in $X \setminus \mathcal{O}$) is normal. In fact the normality of X allows to identify the ring $K[X]$ of regular functions on X with the functions on the orbit $\mathcal{O}$ and so, by Frobenius reciprocity, to analyse $K[X]$ as a representation of G (cf. [11], [1]). In our case this is closely related to the "multiplicity conjecture" of Dixmier; we refer the reader to the paper [1] for a detailed description of this connection and some applications.

A different proof of this theorem will appear in [23].

0.2. The theorem has also another interesting application, shown to us by Th. Vust, in the spirit of the classical theory of Schur. If U is a finite dimensional vector space one has the classical relation between the action of $GL(U)$ and of the symmetric group $\mathfrak{S}_m$ on the tensor space $U^{\otimes m}$. If we restrict to the subgroup G_A of $GL(U)$ centralizing a fixed matrix A ($\in \operatorname{End} U$), then one can still compute the centralizer of G_A acting on $U^{\otimes m}$ and one obtains (see Sect. 6): $\operatorname{End}_{G_A}(U^{\otimes m})$ *is spanned by the endomorphisms*

$$\sigma \cdot A^{h_1} \otimes A^{h_2} \otimes \ldots \otimes A^{h_m} \qquad (\sigma \in \mathfrak{S}_m, h_1, \ldots, h_m \in \mathbb{N}).$$

We remark that the group G_A is not reductive and the commuting algebra is not semisimple in general.

0.3. In many ways the motivation to study this problem came from a fundamental paper of B. Kostant [11] in which he studies in detail the adjoint action on a semisimple Lie algebra $\mathfrak{g}$. In the course of his analysis he shows the normality of

0020-9910/79/0053/0227/$04.20

the variety $\overline{C_A}$ in the case in which A is a *regular nilpotent element* of $\mathfrak{g}$ (i.e. $\overline{C_A}$ is the *nilpotent cone* of $\mathfrak{g}$). His method depends on the fact that, in this case, $\overline{C_A}$ can be proved to be a *complete intersection* in $\mathfrak{g}$. This is no more true for the non regular classes in general, nevertheless some particular cases were treated by W. Hesselink [8]; we wish to thank him for his comments on an earlier version of the paper. Our method, on the other hand, consists in constructing an auxiliary variety Z which is a complete intersection and of which $\overline{C_A}$ is a "quotient" (1.4).

0.4. Remark. It is known (see [8] proposition 1, or use the method of associated cones [1]) that it is sufficient for the proof of the theorem to treat the case of a *nilpotent matrix A* and so we restrict to this case. Then $\overline{C_A}$ has a resolution of singularities $\pi: X \to \overline{C_A}$ where X is the cotangent bundle of GL_n/P, P a parabolic subgroup of GL_n ([4], or [1] Anhang). Then the canonical divisor of X is 0 and so by the theorem of Grauert-Riemenschneider (cf. [9] p. 50) it follows that $\overline{C_A}$ has *rational singularities* and the normality of $\overline{C_A}$ is sufficient to insure also the Cohen-Macaulay property. So the main point of the paper is to prove that $\overline{C_A}$ is *normal.* The proof we give should be adaptable also to positive characteristic; it yields at least that the *normalisation of* $\overline{C_A}$ *is purely inseparable over* $\overline{C_A}$ (cf. remark 5.7).

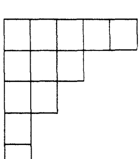
0.5. Let us remark finally that the methods developed here have analogues for all the classical groups. In this case, which will be treated in a subsequent paper, there occur different phenomena which are not yet fully understood. Of course the non connected conjugacy classes have non normal closure, but there are also *infinitely many* connected conjugacy classes C_A for which $\overline{C_A}$ is *not normal*; the simplest known cases are: for the symplectic groups the one in $\mathfrak{sp}_8$ relative to the partition $(3,3,1,1)$, for the orthogonal groups the one in $\mathfrak{so}_{13}$ relative to the partition $(4,4,2,2,1)$.

1. Notations, Some Known Results

1.1. Let us fix some notations. Any nilpotent matrix is conjugate to one in normal Jordan block form:

$$\begin{pmatrix} J_{p_1} & 0 & 0 & \dots & 0 \\ 0 & J_{p_2} & 0 & & \vdots \\ 0 & 0 & & & \vdots \\ 0 & 0 & \dots & \dots & J_{p_k} \end{pmatrix}, \quad J_t := \begin{pmatrix} 0 & 1 & & 0 \\ & 0 & 1 & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix} \quad \text{a } t \times t\text{-block.} \qquad (*)$$

We can assume $p_1 \geqq p_2 \geqq \dots \geqq p_k$; this decreasing sequence $\eta = (p_1, p_2, \dots, p_k)$ is a partition of n and it is convenient to represent it geometrically as a *Young-diagram* with rows consisting of $p_1, p_2, \dots, p_k$ boxes respectively:

e.g. the diagram  corresponds to the partition (5, 3, 2, 1, 1).

The *dual partition* $\hat{\eta}=(\hat{p}_1,\hat{p}_2,\ldots,\hat{p}_m)$ is defined setting $\hat{p}_i$ equal to the length of the i^{th} column of the diagram η; more formally $\hat{p}_i:=\#\{j|p_j\geqq i\}$. In case of a partition η associated to the normal Jordan block form of a nilpotent matrix A, the dual partition $\hat{\eta}$ has the following interpretation:

$$\dim \operatorname{Ker} A^j=\sum_{i=1}^{j}\hat{p}_i$$

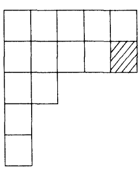
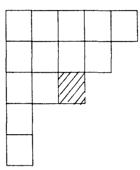
or equivalently

$$\operatorname{rk} A^j=\sum_{i>j}\hat{p}_i.$$

1.2. Given two partitions $\eta=(p_1,\ldots,p_s)$ and $\nu=(q_1,\ldots,q_t)$ of n, *we say* $\eta\geqq\nu$, *if we have* $\sum_{i=1}^{j}p_i\geqq\sum_{i=1}^{j}q_i$ *for all* j. This is equivalent to $\sum_{k>j}\hat{p}_k\geqq\sum_{k>j}\hat{q}_k$ for all j.

A simple property of this ordering, which expresses it geometrically, is the following (cf. [7] Proposition 3.9):

Proposition. *If* $\eta>\nu$ *and no other partition is in between them* (i.e. η and ν are adjacent in the ordering), *then the diagram of* η *is obtained from the one of* ν *raising a box from one row to the first allowable position.*

(e.g.  $=\eta$ and  $=\nu$)

1.3. From now on, if η is a partition of n, we will indicate with C_η the conjugacy class of the matrix (*) in normal Jordan block form with partition η. The following is the basic theorem on degenerations of orbits (cf. [7] Theorem 3.10 and Corollary 3.8 (a)).

Proposition. a) *Given two partitions* η *and* ν *of* n, *we have* $\eta\geqq\nu$ *if and only if* $\overline{C_\eta}\supseteq C_\nu$.

b) *If* $\eta=(p_1,\ldots,p_t)$ *is a partition of* n *and* $\hat{\eta}=(\hat{p}_1,\ldots,\hat{p}_k)$ *the dual partition, we have:*

$$\dim C_\eta=n^2-\sum_{i,j=1}^{t}\min(p_i,p_j)=n^2-\sum_{i=1}^{k}\hat{p}_i^2=2\sum_{i<j}\hat{p}_i\hat{p}_j.$$

1.4. We are working always with affine varieties and we will use the following terminology. If X is an (affine) variety with the action of a reductive group G and $\pi\colon X\to Y$ a morphism, we say that π is a *quotient* (under G), if the coordinate ring of Y is identified, via π, with the ring of *G-invariant functions* on X. We denote this quotient by $\pi\colon X\to X/G$. The following properties of quotient maps are well known ([20], Chap. 1, §2):

a) *Let* $Z\subseteq X$ *be a G-stable closed subvariety. Then* $\pi(Z)\subseteq X/G$ *is closed and* $\pi|_Z$: $Z\to\pi(Z)$ *is a quotient.*

b) *Consider the following fibre product:*

$$\begin{array}{ccc} X' := Y\times_{X/G} X & \xrightarrow{\phi'} & X \\ \Big\downarrow{\scriptstyle \pi'} & & \Big\downarrow{\scriptstyle \pi} \\ Y & \xrightarrow{\phi} & X/G. \end{array}$$

The action of G on X induces an action on the fibre product X' in a natural way and π' is a quotient with respect to this action.

2. The Induction Lemma

2.1. If U, V are vector spaces we will write $\mathrm{L}(U,V)$ for the space of linear maps from U to V and $\mathrm{L}(U)$ instead of $\mathrm{L}(U,U)$. If V is n dimensional and η is a partition of n, we may consider the elements of $\mathrm{L}(V)$ as $n\times n$ matrices and so $C_\eta \subseteq \mathrm{L}(V)$.

2.2. Let $\eta=(p_1,\dots,p_k)$ be a partition of n. Erasing the *first column* in the Young diagram η one obtains a partition $\eta'=(p'_1,p'_2,\dots,p'_k)$ of $m:=n-\hat{p}_1=n-k$, formally defined by $p'_i=p_i-1$ for all i. In terms of dual partitions we have $\hat{\eta}'=(\hat{p}_2,\hat{p}_3,\dots)$.
Fix vector spaces U, V of dimension m, n respectively and consider the two maps

$$\begin{array}{ccc} \mathrm{L}(U,V)\times\mathrm{L}(V,U) & \xrightarrow{\pi} & \mathrm{L}(U) \\ \Big\downarrow{\scriptstyle \rho} & & \\ \mathrm{L}(V) & & \end{array}$$

defined by $\pi(A,B)=BA$, $\rho(A,B)=AB$.

Theorem (First fundamental theorem of invariant theory): *π and ρ are quotient maps* (under $GL(V)$, $GL(U)$ respectively) *and the image of ρ is the determinantel variety of matrices of rank $\leqq m$.* (cf. [22] § 3, Théorème 3 or [18] II.6, Theorem 2.6. A; for a characteristic free proof see [2] § 3.)

2.3. Consider finally the orbits $C_{\eta'}\subseteq \mathrm{L}(U)$, $C_\eta \subseteq \mathrm{L}(V)$ and the variety $N_\eta := \pi^{-1}(\overline{C_{\eta'}})$.

Lemma. $\rho(N_\eta)=\overline{C_\eta}$:

$$\begin{array}{ccc} \pi^{-1}(\overline{C_{\eta'}})=N_\eta & \xrightarrow{\pi} & \overline{C_{\eta'}} \\ \Big\downarrow{\scriptstyle \rho} & & \\ \overline{C_\eta} & & \end{array}$$

Proof. First of all we show that $\rho(N_\eta)\subseteq \overline{C_\eta}$, using repeatedly Proposition 1.3 a). Let $(A,B)\in N_\eta$, i.e. $BA\in\overline{C_{\eta'}}$. To prove that $AB\in\overline{C_\eta}$ we must verify that, for any

$i \geqq 1$, we have $\operatorname{rk}(AB)^i \leqq \sum_{j \geqq i+1} \hat{p}_j$. Now $(AB)^i = A(BA)^{i-1} B$, so $\operatorname{rk}(AB)^i \leqq \operatorname{rk}(BA)^{i-1} \leqq \sum_{j \geqq i} \hat{p}'_j = \sum_{j \geqq i+1} \hat{p}_j$ as desired.

To show that $\rho(N_\eta) = C_\eta$ it is sufficient to prove that $C_\eta \subseteq \rho(N_\eta)$ (since ρ is a quotient map and so $\rho(N_\eta)$ is closed, cf. 1.4a). Let us then fix $D \in C_\eta$, $D\colon V \to V$. We can clearly identify U with $D(V)$ since $\operatorname{rk} D = m$. It is immediate to verify that $D|_U$ has Young diagram η' and clearly we have a factorization

$$\begin{tikzcd} V \arrow[rr, "D"] \arrow[dr, "B"'] & & V \\ & U \arrow[ur, "A"'] & \end{tikzcd}$$

where A is the inclusion and B coincides with D. On the other hand BA is just $D|_U$ so that the pair (A, B) is in N_η and the claim is proved[1]. qed.

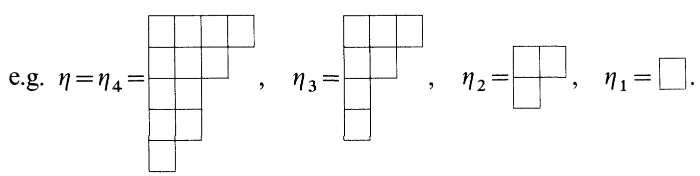
We will use this lemma to present the variety $\overline{C_\eta}$ as a quotient of a suitable variety Z for which we will be able to prove normality (Theorem 3.3).

3. The Variety Z

3.1. Notations being as in section 2 we make the following construction. Starting with a fixed partition $\eta = (p_1, \ldots, p_k)$ and dual partition $\hat{\eta} = (\hat{p}_1, \ldots, \hat{p}_t)$, $t := p_1$, we define a sequence of partitions

$$\eta_t := \eta, \quad \eta_{t-1}, \quad \eta_{t-2}, \ldots, \eta_1$$

by $\eta_{i-1} := \eta'_i$ (i.e. by erasing successively the first column of the corresponding Young diagrams);



Then η_i is a partition of $n_i := \hat{p}_t + \hat{p}_{t-1} + \ldots + \hat{p}_{t-i+1}$ with dual partition $\hat{\eta}_i = (\hat{p}_{t-i+1}, \ldots, \hat{p}_{t-1}, \hat{p}_t)$.

Construct next vector spaces $U_1, U_2, \ldots, U_t$ of dimensions $n_1, n_2, \ldots, n_t$ respectively and consider the affine spaces:

$$\begin{aligned} M := {} & \mathrm{L}(U_1, U_2) \times \mathrm{L}(U_2, U_1) \times \mathrm{L}(U_2, U_3) \\ & \times \mathrm{L}(U_3, U_2) \times \ldots \times \mathrm{L}(U_{t-1}, U_t) \times \mathrm{L}(U_t, U_{t-1}) \end{aligned}$$

and

$$N := \mathrm{L}(U_1) \times \mathrm{L}(U_2) \times \ldots \times \mathrm{L}(U_{t-1}).$$

[1] This argument due to W. Hesselink replaces a more direct matrix computation we had made.

We will indicate a point α of M by

$$\alpha=(A_1, B_1, A_2, B_2, \ldots, A_{t-1}, B_{t-1})$$

where $A_i\colon U_i \to U_{i+1}$, $B_i\colon U_{i+1} \to U_i$.

3.2. We are now ready to define the variety Z. *It is the subvariety of* M *defined by the equations*

$$\begin{aligned} &B_1 A_1=0\\ &B_2 A_2=A_1 B_1\\ &B_3 A_3=A_2 B_2\\ &\quad\vdots\\ &B_{t-1} A_{t-1}=A_{t-2} B_{t-2}. \end{aligned} \tag{$**$}$$

In more suggestive notation we write

$$\alpha\colon U_0=0 \underset{B_0=0}{\overset{A_0=0}{\rightleftarrows}} U_1 \underset{B_1}{\overset{A_1}{\rightleftarrows}} U_2 \underset{B_2}{\overset{A_2}{\rightleftarrows}} U_3 \ldots U_{t-1} \underset{B_{t-1}}{\overset{A_{t-1}}{\rightleftarrows}} U_t$$

for the elements of Z. The equations just require that for each $i=1 \ldots t-1$ the two compositions $U_{i-1} \rightleftarrows U_i \rightleftarrows U_{i+1}$ yield the same endomorphism of U_i. (See also [19] 5.3, where this objects occur as representations of a certain Lie algebra.)

In due time we will prove that the equations we have given actually define a reduced variety; for the moment we think of Z as a scheme, possibly not reduced, and we indicate by Z_{red} the reduced variety associated.

The best way to understand the equations is to construct a map $\Phi\colon M \to N$ given by the formula:

$$\begin{aligned} &\Phi(A_1, B_1, A_2, B_2, \ldots, A_{t-1}, B_{t-1})\\ &\quad=(B_1 A_1, B_2 A_2-A_1 B_1, B_3 A_3-A_2 B_2, \ldots, B_{t-1} A_{t-1}-A_{t-2} B_{t-2}). \end{aligned}$$

Then Z, as a scheme, is the fiber $\Phi^{-1}(0)$ of the 0 point in N.

3.3. Consider the group $G:=GL(U_1) \times GL(U_2) \times \ldots \times GL(U_t)$ and its normal subgroup $H:=GL(U_1) \times GL(U_2) \times \ldots \times GL(U_{t-1})$. The group G acts on M and N in a natural way:

On

$$\begin{aligned} &M\colon (g_1, g_2, \ldots, g_t)(A_1, B_1, A_2, B_2, \ldots, A_{t-1}, B_{t-1})\\ &\quad:=(g_2 A_1 g_1^{-1}, g_1 B_1 g_2^{-1}, \ldots, g_t A_{t-1} g_{t-1}^{-1}, g_{t-1} B_{t-1} g_t^{-1}) \end{aligned}$$

and on

$$\begin{aligned} &N\colon (g_1, g_2, \ldots, g_t)(E_1, E_2, \ldots, E_{t-1})\\ &\quad:=(g_1 E_1 g_1^{-1}, \ldots, g_{t-1} E_{t-1} g_{t-1}^{-1}). \end{aligned}$$

It is easy to verify that the map Φ: $M \to N$ is equivariant under G and so Z is invariant under G. Now the main theorem is an immediate consequence of the following more precise result (use the fact that a quotient of a normal variety is also normal):

Theorem. i) *Z is a complete intersection in M; the equations* (**) *give a regular sequence.*

ii) *Z is non singular in codimension* 1.

iii) *Z is reduced, irreducible and normal.*

iv) *There is an isomorphism $Z/H \xrightarrow{\sim} \overline{C_\eta}$ being compatible with the actions of $GL(U_t) = G/H$.*

The rest of the paper is devoted to the proof of this theorem. We first reduce it to a lemma (3.7) whose proof will be given in Sect. 5 (5.4, 5.5, 5.6) using the theory of nilpotent pairs (section 4) and a dimension formula for nilpotent pair orbits (5.3).

3.4. First of all we settle part iv) which gives the connection between Z and $\overline{C_\eta}$. We consider the map

$$\Theta: M \to \mathrm{L}(U_t)$$

given by $(A_1, B_1, \dots, A_{t-1}, B_{t-1}) \mapsto A_{t-1} B_{t-1}$, which is clearly $GL(U_t)$ equivariant.

Proposition. *$\Theta(Z_{\mathrm{red}}) = \overline{C_\eta}$ and the induced map Θ': $Z_{\mathrm{red}} \to \overline{C_\eta}$ is a quotient map under H* (i.e. $Z_{\mathrm{red}}/H \xrightarrow{\sim} \overline{C_\eta}$).

Proof. We use repeatedly lemma 2.3. Since $B_1 A_1 = 0$ the pair (A_1, B_1) is in the variety N_{η_2} and so $A_1 B_1 \in \overline{C_{\eta_2}}$. By induction we may assume $A_{i-1} B_{i-1} \in \overline{C_{\eta_i}}$. Since $\underline{B_i A_i} = A_{i-1} B_{i-1}$ we have that $(A_i, B_i) = N_{\underline{\eta_i}+1}$ and so again by 2.3, $A_i B_i \in \overline{C_{\eta_{i+1}}}$. Thus finally Θ maps Z_{red} into $\overline{C_\eta} = \overline{C_{\eta_t}}$ and the same lemma 2.3 applied inductively shows that Z_{red} is mapped onto $\overline{C_\eta}$. To see that the map is a quotient under H we perform the quotients in succession. First under $GL(U_1)$, we have the quotient map (Theorem 2.2)

$$\Theta_1: M \to \mathrm{L}(U_2) \times \mathrm{L}(U_2, U_3) \times \mathrm{L}(U_3, U_2) \times \dots \\ \times \mathrm{L}(U_{t-1}, U_t) \times \mathrm{L}(U_{t-1}\, U_{t-1})$$

given by

$$(A_1, B_1, A_2, B_2, \dots, A_{t-1}, B_{t-1}) \mapsto (A_1 B_1, A_2, B_2, \dots, A_{t-1}, B_{t-1}).$$

If we restrict this map to Z_{red} we have again a quotient map (since we are in characteristic zero cf. 1.4a). Now on Z_{red} we have $A_1 B_1 = B_2 A_2$ (if $t > 2$); thus we see that Θ_1 maps Z_{red} into the graph of the map

$$\gamma: M_1 := \mathrm{L}(U_2, U_3) \times \mathrm{L}(U_3, U_2) \times \dots \\ \times \mathrm{L}(U_{t-1}, U_t) \times \mathrm{L}(U_t, U_{t-1}) \to \mathrm{L}(U_2) \\ (A_2, B_2, A_3, B_3, \dots, A_{t-1}, B_{t-1}) \mapsto B_2 A_2.$$

Thus we may replace the graph of γ with its domain and drop the first coordinate $A_1 B_1$. Hence on Z_{red} the mapping

$$(A_1, B_1, \ldots, A_{t-1}, B_{t-1}) \mapsto (A_2, B_2, \ldots, A_{t-1}, B_{t-1}) \in M_1$$

is a quotient under $GL(U_1)$. Its image $Z_1 \subseteq M_1$ is easily seen to be defined by the "equations":

$$\begin{aligned} &B_2 A_2 \in \overline{C_{\eta_2}} \\ &A_2 B_2 = B_3 A_3 \\ &A_3 B_3 = B_4 A_4 \\ &\quad\vdots \\ &A_{t-2} B_{t-2} = B_{t-1} A_{t-1}. \end{aligned}$$

Similarly (if $t>3$) $Z_1/GL(U_2) = Z_{\text{red}}/GL(U_1) \times GL(U_2)$ is naturally contained in

$$M_2 := \mathrm{L}(U_3, U_4) \times \mathrm{L}(U_4, U_3) \times \ldots \times \mathrm{L}(U_{t-1}, U_t) \times \mathrm{L}(U_t, U_{t-1})$$

and given by the "equations":

$$\begin{aligned} &B_3 A_3 \in \overline{C_{\eta_3}} \\ &A_3 B_3 = B_4 A_4 \\ &\quad\vdots \\ &A_{t-2} B_{t-2} = B_{t-1} A_{t-1}. \end{aligned}$$

Then finally by induction $Z_{\text{red}}/GL(U_1) \times GL(U_2) \times \ldots \times GL(U_{t-2})$ is given by

$$\{(A_{t-1}, B_{t-1}) \mid B_{t-1} A_{t-1} \in \overline{C_{\eta_{t-1}}}\} \subseteq \mathrm{L}(U_{t-1}, U_t) \times \mathrm{L}(U_t, U_{t-1}),$$

i.e. it is the variety $N_\eta = N_{\eta_t}$, and hence

$$Z_{\text{red}}/H \cong N_\eta / GL(U_{t-1}) \cong \overline{C_\eta}$$

(the isomorphism being induced by Θ). qed.

This reasoning can be also displayed in a more suggestive way constructing a diagram, e.g. $t=5$:

$$\begin{tikzcd}
Z(1,5) \arrow[r] \arrow[d] & Z(1,4) \arrow[r] \arrow[d] & Z(1,3) \arrow[r] \arrow[d] & N_{\eta_1} \arrow[r] \arrow[d] & \overline{C_{\eta_1}} \\
Z(2,5) \arrow[r] \arrow[d] & Z(2,4) \arrow[r] \arrow[d] & N_{\eta_2} \arrow[r] \arrow[d] & \overline{C_{\eta_2}} & \\
Z(3,5) \arrow[r] \arrow[d] & N_{\eta_3} \arrow[r] \arrow[d] & \overline{C_{\eta_3}} & & \\
N_{\eta_4} \arrow[r] \arrow[d] & \overline{C_{\eta_4}} & & & \\
\overline{C_{\eta_5}} & & & &
\end{tikzcd}$$

where each $Z(i,j)$ is constructed inductively forming a fiber product. If we proceed on a column we see that

$$Z(k,n) \xleftarrow{\sim} Z(1,n)/GL(U_1)\times GL(U_2)\times \ldots \times GL(U_{k-1}) \quad \text{for } k<n-1$$

(cf. 1.4b).

3.5. We now make a simple remark on the basic map Φ: $M \to N$ (3.2) for which $Z=\Phi^{-1}(0)$.

Let M^0 be the open subset of M of those elements $(A_1, B_1, A_2, B_2, \ldots, A_{t-1}, B_{t-1})$ such that *for each* $i=1,2,\ldots,t-1$ *either* A_i *or* B_i *has maximal rank.* Then we have:

Proposition. *The differential* $d\Phi$ *of* Φ *is onto at every point* α *of* M^0.

Proof. Let $\alpha=(A_1, B_1, A_2, B_2, \ldots, A_{t-1}, B_{t-1})\in M^0$. We can identify the tangent space of M in α with M itself and take a point $T=(X_1, Y_1, X_2, Y_2, \ldots, X_{t-1}, Y_{t-1})$ in it. Then the tangent map gives

$$\begin{aligned} d\Phi_\alpha(T) = (&Y_1 A_1 + B_1 X_1, Y_2 A_2 + B_2 X_2 - X_1 B_1 - A_1 Y_1, \ldots, Y_{t-1} A_{t-1} \\ &+ B_{t-1} X_{t-1} - X_{t-2} B_{t-2} - A_{t-2} Y_{t-2}). \end{aligned}$$

If $W=(W_1, W_2, \ldots, W_{t-1})$ is any tangent vector in $\Phi(\alpha)\in N$ we can solve inductively the equations

$$\begin{aligned} &Y_1 A_1 + B_1 X_1 = W_1 \\ &Y_2 A_2 + B_2 X_2 - X_1 B_1 - A_1 Y_1 = W_2 \\ &\vdots \\ &Y_{t-1} A_{t-1} + B_{t-1} X_{t-1} - X_{t-2} B_{t-2} - A_{t-2} Y_{t-2} = W_{t-1} \end{aligned}$$

provided that for each i either A_i or B_i has maximal rank. In fact if A_i has maximal rank then there is an $\overline{A}_i$: $U_{i+1} \to U_i$ with $\overline{A}_i A_i = Id_{U_i}$, so the equation $Y_i A_i = R_i$ is solved by $Y_i := R_i \overline{A}_i$. Similarly if B_i has maximal rank then there is an element $\overline{B}_i$: $U_i \to U_{i+1}$ with $B_i \overline{B}_i = Id_{U_i}$ and the equation $B_i X_i = S_i$ is solved by $X_i := \overline{B}_i S_i$. qed.

3.6. The net result of this proposition is this:

Corollary. *The open subvariety* $Z^0 := Z \cap M^0$ *of* Z *is smooth and of codimension* $\sum_{i=1}^{t-1} n_i^2$ *in* M.

Proof. The only thing to prove is that $Z^0 \neq \emptyset$ since then the statement is a consequence of 3.5 $\left(\dim N = \sum_{i=1}^{t-1} n_i^2 \text{ by definition 3.1}\right)$. Now if we recall the proof of 3.4 we see that we have constructed an element in Z such that for all i both A_i and B_i have maximal rank (see also the construction in 3.1). More precisely if D: $U \to U$ is any element of C_η we may assume $U_t = U$, $U_{t-1} = D(U)$, $U_{t-2} = D^2(U), \ldots, U_1 = D^{t-1}(U)$, $U_0 = 0$. Setting A_i: $U_i \to U_{i+1}$ the inclusion, and B_i: $U_{i+1} \to U_i$ the map D itself, we have the required element. qed.

Remark. If we insist that for each i both A_i and B_i have maximal rank we still get a non empty open set Z' of Z. One can easily show that Z' is an orbit under G (cf. proof above). It will be proved in fact that this is the unique open orbit of G in Z (5.4).

3.7. To complete the proof of the theorem it is enough to show the following result:

Lemma. $\dim(Z \setminus Z^0) \leqq \dim Z - 2$.

In fact using 3.6 this lemma implies $\dim Z = \dim Z^0$ and that Z is non singular in codimension one. Thus again by 3.6 we have that the codimension of Z in M is exactly the number $\sum_{i=1}^{t-1} n_i^2$ of equations defining Z (3.2). This implies that these equations form a regular sequence and hence Z is a complete intersection. Since Z is a cone it is also connected. But then by Serre's criterion ([6] IV, Théorème 5.8.6) Z is normal reduced and so also irreducible. This completes the proof of the theorem 3.3 modulo the lemma above.

For this basic statement we will need to stratify the complement of Z^0 in Z with strata of which we can compute the dimension (5.1, 5.3) and this will lead us to the theory of nilpotent pairs (cf. the following section).

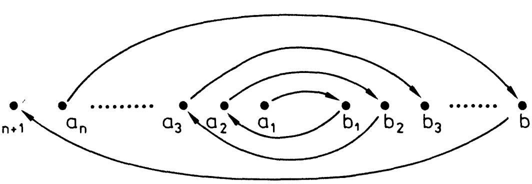
4. Nilpotent Pairs

Given two vector spaces U, V we consider the space $L := \mathrm{L}(U, V) \times \mathrm{L}(V, U)$ of pairs of maps $U \underset{B}{\overset{A}{\rightleftarrows}} V$ as a representation of $GL(U) \times GL(V)$ in a canonical way:

$$(X, Y)(A, B) = (YAX^{-1}, XBY^{-1}).$$

The theory of orbits for this representation is known (cf. [3], [14], or [5]) and it is in fact a special case of the theory of vector space crowns. One can naturally think of such pairs as of a category of modules, and the classification is (like in the case of Jordan blocks) through indecomposable modules. Also the "invariant theory" of this representation is well known (see [12] and also [17]). In our case we are interested in a special class of pairs, those $U \underset{B}{\overset{A}{\rightleftarrows}} V$ for which BA (or equivalently AB) is *nilpotent*. We will call such pairs "*nilpotent pairs*". They can be easily seen to be exactly the unstable vectors (in the sense of geometric invariant theory) of the representation L.

4.2. The classification of the *indecomposable nilpotent pairs* is rather simple and resembles the theory of Jordan blocks. The indecomposables are of the following types:

M_{2n+1}:


i.e. the space U is spanned by the basis $a_1, a_2, \ldots, a_{n+1}$, V has basis $b_1, b_2, \ldots, b_n$ and

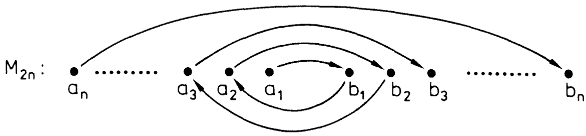
$$A\,a_i = b_i, \qquad B\,b_j = a_{j+1}.$$

This type will be in short indicated by a string

abababab...ba

with $n+1$ a's and n b's.

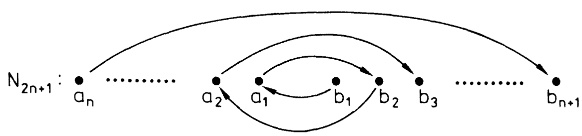
The type

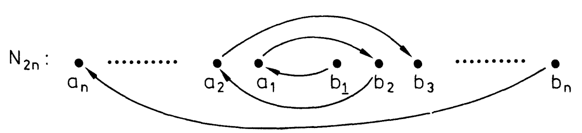


is defined in a similar way and is shortly indicated by the string

abab...ab

with n a's and n b's.
We have two other types starting with b instead of a:





shortly indicated by *baba...bab* and *baba...ba* respectively.

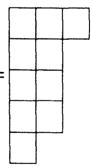
4.3. In general a nilpotent pair $U \underset{B}{\overset{A}{\rightleftarrows}} V$ is a direct sum of indecomposables and so it will be determined by a finite set of such strings (*ab-strings*). We will refer to such a set of strings as the *ab-diagram of the pair*. It is easy to see that two distinct *ab*-diagrams give rise to non isomorphic pairs, since one can easily recover the *ab*-diagram from the knowledge of the ranks of all the compositions $BABA\ldots$.

Given a nilpotent pair $U \underset{B}{\overset{A}{\rightleftarrows}} V$ through its *ab*-diagram δ, it is simple to recognise the Young diagrams of the nilpotent matrices $BA\colon U \to U$ and $AB\colon V \to V$: For the diagram of BA *suppress all the b's in the ab-strings of δ*. In this way every *ab*-string gives rise to a string of a's which can be interpreted as a row in a Young diagram. Similarly for AB one has to suppress all the a's. We call these diagrams the *associated a-diagram* and the *associated b-diagram* of δ and denote them as in 2.2 by $\pi(\delta)$ and $\rho(\delta)$.

$$\text{e.g. } \delta := \begin{array}{l} a\,b\,a\,b\,a\,b \\ b\,a\,b\,a \\ a\,b\,a \\ a\,b\,a \\ b\,a \\ b \end{array}$$

then

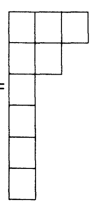
$$\pi(\delta) = \begin{array}{l} a\,a\,a \\ a\,a \\ a\,a \\ a\,a \\ a \end{array} =$$



$= (3, 2, 2, 2, 1)$

is the Young diagram of $BA \in \mathrm{L}(U)$ and

$$\rho(\delta) = \begin{array}{l} b\,b\,b \\ b\,b \\ b \\ b \\ b \\ b \end{array} =$$



$= (3, 2, 1, 1, 1, 1)$

is that of $AB \in \mathrm{L}(V)$.

4.4. One should make three remarks:

Remark 1. Not all pairs of Young diagrams describing a nilpotent orbit in $\mathrm{L}(U)$ and one in $\mathrm{L}(V)$ are associated to some nilpotent pair. Furthermore, there can be different nilpotent pairs giving rise to the same pair of Young diagrams.

Remark 2. For a nilpotent pair (A, B) with ab-diagram δ one can immediately verify the following:

i) *A is injective if and only if every ab-string in δ ends with b.*
ii) *A is surjective if and only if every ab-string in δ starts with a.*
iii) *B is injective if and only if every ab-string in δ ends with a.*
iv) *B is surjective if and only if every ab-string in δ starts with b.*

Remark 3. For any ab-diagram δ we denote by X_δ its orbit in L (under the group $GL(U) \times GL(V)$). We have the two maps

$$\begin{tikzcd} X_\delta \arrow[r, "\pi'"] \arrow[d, "\rho'"'] & C_{\pi(\delta)} \\ C_{\rho(\delta)} & \end{tikzcd}$$

induced by π and ρ (2.2) which are *fibrations* (being of the form $G/H \to G/H'$ with closed subgroups $H \subseteq H' \subseteq G := GL(U) \times GL(V)$). In particular π' and ρ' are *smooth.*

5. Nilpotent Strings, Proof of Lemma 3.7

5.1. We want to go back to the basic variety Z (3.2) formed by strings

$$\alpha\colon\ U_0 = 0 \underset{B_0=0}{\overset{A_0=0}{\rightleftarrows}} U_1 \underset{B_1}{\overset{A_1}{\rightleftarrows}} U_2 \underset{B_2}{\overset{A_2}{\rightleftarrows}} U_3 \dots U_{t-1} \underset{B_{t-1}}{\overset{A_{t-1}}{\rightleftarrows}} U_t$$

with the conditions $B_{i+1}A_{i+1} = A_i B_i$ for $i = 0, 1, \dots, t-2$. Let us indicate by $Y_0 = \{\emptyset\}$, $Y_1, Y_2, \dots, Y_t$ the (finite) sets of diagrams indexing nilpotent conjugacy classes in $\mathrm{L}(U_0), \mathrm{L}(U_1), \dots, \mathrm{L}(U_t)$ respectively and by $\Psi_0, \Psi_1, \dots, \Psi_{t-1}$ the (finite) sets of *ab*-diagrams indexing conjugacy classes of nilpotent pairs in

$$\mathrm{L}(U_0, U_1) \times \mathrm{L}(U_1, U_0), \mathrm{L}(U_1, U_2) \times \mathrm{L}(U_2, U_1), \dots,$$
$$\mathrm{L}(U_{t-1}, U_t) \times \mathrm{L}(U_t, U_{t-1})$$

respectively. We have the already described maps associated to the two compositions (cf. 4.3):

$$Y_0 \xleftarrow{\pi_0} \Psi_0 \xrightarrow{\rho_0} Y_1 \xleftarrow{\pi_1} \Psi_1 \xrightarrow{\rho_1} Y_2$$
$$\xleftarrow{\pi_2} \Psi_2 \xrightarrow{\rho_2} Y_3 \leftarrow \dots \to \Psi_{t-1} \xrightarrow{\rho_{t-1}} Y_t.$$

We can form the iterated fiber products and construct the finite set Λ of strings $\lambda = (\delta_0, \delta_1, \delta_2, \dots, \delta_{t-1})$ of *ab*-diagrams $\delta_i \in \Psi_i$ with

$$\rho_i(\delta_i) = \pi_{i+1}(\delta_{i+1}), \qquad i = 0, 1, \dots, t-2.$$

For each $\lambda \in \Lambda$ we have a stratum Z_λ of the variety Z: *Z_λ is the set of all points*

$$\alpha\colon\ 0 = U_0 \underset{B_0}{\overset{A_0}{\rightleftarrows}} U_1 \underset{B_1}{\overset{A_1}{\rightleftarrows}} U_2 \rightleftarrows \dots \rightleftarrows U_{t-1} \underset{B_{t-1}}{\overset{A_{t-1}}{\rightleftarrows}} U_t$$

of Z such that for each i the nilpotent pair (A_i, B_i) has ab-diagram δ_i.

Put $\sigma_i := \rho_{i-1}(\delta_{i-1})$, $i = 1, 2, \dots, t$ and let us indicate as usual by C_{σ_i} the conjugacy class of diagram σ_i and by X_{δ_i} the nilpotent pair orbit of diagram δ_i. The definition of Λ implies that we have a fiber product diagram subordinate to the basic diagram constructing Z:

$$\begin{tikzcd}
Z_\lambda = Z_\lambda(1,t) \arrow[r] \arrow[d] & \dots \arrow[r] & Z_\lambda(1,4) \arrow[r] \arrow[d] & Z_\lambda(1,3) \arrow[r] \arrow[d] & X_{\delta_1} \arrow[r] \arrow[d] & C_{\sigma_1} = (0) \\
Z_\lambda(2,t) \arrow[r] \arrow[d] & \dots \arrow[r] & Z_\lambda(2,4) \arrow[r] \arrow[d] & X_{\delta_2} \arrow[r] \arrow[d] & C_{\sigma_2} & \\
\vdots \arrow[d] & & X_{\delta_3} \arrow[r] \arrow[d] & C_{\sigma_3} & & \\
X_{\delta_t} \arrow[r] \arrow[d] & C_{\sigma_{t-1}} & C_{\sigma_4} \arrow[l, dotted, no head] & & & \\
C_{\sigma_t} & & & & &
\end{tikzcd}$$

in which each map is smooth (4.4 remark 3). Hence we get the following proposition:

Proposition. (i) Z_λ *Is a locally closed, G-stable, smooth and irreducible subvariety of Z.*

(ii) *The set Λ indexes a stratification of Z into smooth G-stable strata.*

5.2. The following result now clearly implies lemma 3.7.

Lemma. *For all $\lambda \in \Lambda$ either $Z_\lambda \subseteq Z^0$ or* $\dim Z_\lambda \leqq \dim Z - 2$.

The proof will be given in 5.4, 5.5, 5.6 using the following dimension formula for nilpotent pair orbits.

5.3. Proposition. *Let $X = X_\delta \subset \mathrm{L}(U, V) \times \mathrm{L}(V, U)$ be a nilpotent pair orbit projecting to the nilpotent conjugacy classes $C_1 \subset \mathrm{L}(U)$ and $C_2 \subset \mathrm{L}(V)$. Then*

$$\dim X_\delta = \tfrac{1}{2}(\dim C_1 + \dim C_2) + \dim U \cdot \dim V - \Delta,$$

$$\Delta := \sum_{i \,\mathrm{odd}} a_i b_i$$

where a_i (resp. b_i) *denotes the number of ab-strings of length i starting with a* (resp. *with b*).

Proof. The representation of $GL(U) \times GL(V)$ on $\mathrm{L} := \mathrm{L}(U, V) \times \mathrm{L}(V, U)$ is a Θ-group in the sence of Vinberg [17] (cf. also [12]): Consider the automorphism Θ of $\mathrm{End}(U \oplus V)$ (and of $GL(U \oplus V)$) given by conjugation with $J = \begin{pmatrix} Id_U & 0 \\ 0 & -Id_V \end{pmatrix}$; then $GL(U) \times GL(V)$ is the fixed point group and $\mathrm{L} \subset \mathrm{End}(U \oplus V)$ the (-1)-

eigenspace of Θ. Furthermore we have the following relation between the dimension of X and the dimension of the conjugacy class $C \subset \operatorname{End}(U \oplus V)$ generated by X:

$$\dim X = \tfrac{1}{2} \dim C$$

(cf. [17] §2.5 Proposition 5, or [12] 1.3 Proposition 5). In order to calculate $\dim C$ denote by r_i resp. s_i the number of a's resp. b's in the i^{th} row of the ab-diagram δ associated to X. Then the partition of the nilpotent conjugacy class C is given by $(p_1, p_2, \ldots)$, $p_i := r_i + s_i$, hence

$$\dim C = (n+m)^2 - \sum_{i,j} \min(p_i, p_j)$$

(1.3 Proposition b), $n := \dim V$, $m := \dim U$). By definition we have $|r_i - s_i| \leqq 1$ and therefore $\min(p_i, p_j) = \min(r_i, r_j) + \min(s_i, s_j)$ except in the case $p_i = p_j$ odd, $r_i = s_j$ and $r_j = s_i$, where $\min(p_i, p_j) = \min(r_i, r_j) + \min(s_i, s_j) + 1$. This implies

$$\begin{aligned} \dim C &= (n+m)^2 - \sum_{i,j} \min(r_i, r_j) + \sum_{i,j} \min(s_i, s_j) + 2 \cdot \sum_{i\,\text{odd}} a_i b_i \\ &= \dim C_1 + \dim C_2 + 2nm - 2\Delta, \end{aligned}$$

hence the required dimension formula.[2] qed.

Now let $\lambda \in \Delta$, $\lambda = (\delta_0, \ldots, \delta_{t-1})$ and $\lambda' = (\delta_0, \ldots, \delta_{t-2})$. Set $\sigma = \rho_{t-1}(\delta_{t-1})$, $\sigma' = \rho_{t-2}(\delta_{t-2})$, and $\Delta_\lambda = \sum_{i=0}^{t-1} \Delta_i$, Δ_i the Δ associated to δ_i.

Corollary. $\dim Z_\lambda = \sum_{i=1}^{t-1} n_i n_{i+1} + \tfrac{1}{2} \dim C_\sigma - \Delta_\lambda$.

Proof. We have the fibre product diagram

$$\begin{tikzcd} Z_\lambda \arrow[r] \arrow[d] & Z_{\lambda'} \arrow[d] \\ X_{\delta_{t-1}} \arrow[r] \arrow[d] & C_{\sigma'} \\ C_\sigma & \end{tikzcd}$$

Now the proposition implies

$$\dim X_{\delta_{t-1}} = \tfrac{1}{2}(\dim C_\sigma + \dim C_{\sigma'}) + \dim U_t \cdot \dim U_{t-1} - \Delta_{t-1},$$

so we have by induction:

$$\begin{aligned} \dim Z_\lambda &= \dim Z_{\lambda'} + \dim X_{\delta_{t-1}} - \dim C_{\sigma'} \\ &= \sum_{i=1}^{t-2} n_i \cdot n_{i+1} + \tfrac{1}{2} \dim C_{\sigma'} + \tfrac{1}{2}(\dim C_{\sigma'} + \dim C_\sigma) \end{aligned}$$

[2] This proof was suggested by G. Kempken; it replaces a explicit but lengthy calculation of stabilizers we have made.

$$+n_{t-1}n_t - \dim C_{\sigma'} - \Delta_{t-1}$$

$$= \sum_{i=1}^{t-1} n_i n_{i+1} + \tfrac{1}{2}\dim C_\sigma - \Delta_\lambda. \quad \text{qed.}$$

5.4. We now look, in view of corollary 5.3, at the projection $\Theta\colon Z \to \overline{C_\eta}$ (3.4) and try to study the various strata Z_λ which lie on top of a given orbit C_σ in $\overline{C_\eta}$. First of all analyze the open orbit C_η: we have to describe the strings $\lambda = (\delta_0, \delta_1, \ldots, \delta_{t-1})$ which lead to $\rho_{t-1}(\delta_{t-1}) = \eta$. We claim that *there is only one such string.*

Let us take in general a Young diagram σ and let σ' be the diagram obtained from σ erasing the first column. We want to find an ab-diagram such that the a-diagram and b-diagram associated are σ' and σ.

Given σ (as b-diagram) and a certain number m of a's, to construct an ab-diagram over σ one has to proceed as follows: First of all every b-string of σ has to be filled internally with a's. This requires altogether as many a's as the number of boxes m' in σ'. If m is equal to m' the ab-diagram δ over σ is unique, its associated a-diagram is σ', and every ab-string of δ starts and ends with b. If $m < m'$ there is no ab-diagram over σ. If $m > m'$, after having used the m' a's, one can utilize the remaining $m - m'$ a's in many ways: add an a at the beginning or the end of an ab-string or create a row with single a.

$$\text{Example:}\quad \sigma = \begin{matrix} b\,b\,b\,b\,b \\ b\,b\,b \\ b\,b\,b \\ b\,b \\ b \end{matrix} = (5,3,3,2,1),\ m' = 9;$$

the unique ab-diagram with $m = 9$ a's is

$$\delta = \begin{matrix} b\,a\,b\,a\,b\,a\,b\,a\,b \\ b\,a\,b\,a\,b \\ b\,a\,b\,a\,b \\ b\,a\,b \\ b \end{matrix}$$

associated to σ and

$$\pi(\delta) = \begin{matrix} a\,a\,a\,a \\ a\,a \\ a\,a \\ a \end{matrix} = \sigma'$$

If we give $m = 10$ a's (for instance) one easily sees that it is possible to construct 11 ab-diagrams over σ, if $m = 11$, we can construct 56 ab-diagrams etc.

Summing up the result we see by induction that *there is a unique string* $\lambda^0 = (\delta_1^0, \delta_2^0, \ldots, \delta_{t-1}^0)$ *such that* $\rho_{t-1}(\delta_{t-1}^0) = \eta = \eta_t$. For this string we have $\rho_i(\delta_i^0) = \eta_{i+1}$ for all i. Since every ab-string in each δ_i^0 starts and ends with b, it follows from 4.4 remark 2 that $Z_{\lambda^0} \subseteq Z^0$ ($Z_{\lambda^0} = Z'$ with the notations of remark 3.6). Hence we get the following result:

Lemma. *There is a unique string* $\lambda^0 = (\delta_1^0, \delta_2^0, \ldots, \delta_{t-1}^0)$ *such that* $\rho_{t-1}(\delta_{t-1}^0) = \eta = \eta_t$. *For this string we have* $Z_{\lambda^0} \subseteq Z^0$.

5.5. We now make a further simple remark. By 3.6 and 3.1 we have

$$\dim Z^0 = 2\sum_{i=1}^{t-1} n_i n_{i+1} - \sum_{i=1}^{t-1} n_i^2 = \sum_{i=1}^{t-1} n_i n_{i+1} + \sum_{i=1}^{t-1} n_i(n_{i+1} - n_i)$$
$$= \sum_{i=1}^{t-1} n_i n_{i+1} + \sum_{i<j} \hat{p}_i \hat{p}_j = \sum_{i=1}^{t-1} n_i n_{i+1} + \tfrac{1}{2} \dim C_\eta .$$

Now if $\lambda \neq \lambda^0$, Z_λ projects to some orbit C_σ with $\sigma < \eta$ and thus $\dim Z_\lambda < \dim Z^0$ (Corollary 5.3). This implies (as one can also verify directly) that $\dim Z_{\lambda^0} = \dim Z^0$ and Z_{λ^0} is the unique open orbit of G acting on Z. The same estimate shows that, if $\dim C_\sigma \leqq \dim C_\eta - 4$, then $\dim Z_\lambda \leqq \dim Z^0 - 2$.

To complete the proof of 5.2 we are thus restricted to analyze the strings λ such that Z_λ projects to some C_σ with $\dim C_\sigma = \dim C_\eta - 2$. In each degeneration the dimension of a nilpotent orbit decreases by at least 2 (since the orbits are even dimensional, see proposition 1.3b). Thus the only case in which we may have $\dim C_\sigma = \dim C_\eta - 2$ is if the diagram σ is obtained from η moving down a single box (Proposition 1.2). The explicit dimension formula (Proposition 1.3b)) shows, in fact, that the only case is *to move a box down to the next row*. Given such a σ we must study which strings $(\delta_1, \delta_2, \ldots, \delta_{t-1})$ lead to $\rho_{t-1}(\delta_{t-1}) = \sigma$.

5.6. We analyze inductively this problem as before and claim that the analysis restricts to the following problem:

Given a diagram η, *let* η' *be obtained from* η *removing the first column, and let* σ *be a one step degeneration of* η (obtained by moving down one box to the next row). *We must study the ab-diagrams* δ *such that* $\rho(\delta) = \sigma$ *and* $\pi(\delta) \leqq \eta'$ ($\pi(\delta)$ and η' with the same number of boxes).

We follow the same analysis as before, letting m be the number of boxes of η'. Let σ' be obtained from σ erasing the first column and m' its number of boxes. *We have clearly the two possibilities* $m = m'$ *and* $m = m' + 1$.

Case I: $m = m'$; this case is obtained when *the box moved in the degeneration of* η *to* σ *is attached to a non empty row.*

In this case the previous analysis (5.4) shows that i) there is a unique *ab*-diagram δ over σ, ii) δ is associated to σ, σ' and σ' is a one step degeneration of η', and iii) every *ab*-string of δ starts and ends with b.

Case II: $m = m' + 1$; this case is obtained when σ *is gotten from* η *splitting off one box from the last row* (to form a new "one box" row). In this case the previous analysis shows that, to form an *ab*-diagram δ over σ, we are forced to place m' a's; the remaining single a can be placed only in the last two rows or by itself since we must preserve the condition $\pi(\delta) \leqq \eta'$.

Let us consider thus the last two rows; after filling with m' of the a's they are:

$b\,a\,b\,a\,b \ldots a\,b$
b

For the remaining a we have 5 choices:

(α) and (α'): *We attach a to the row bab...ab either to the right or to the left.* In this case the ab-diagram δ represents a pair in which one of the two maps has maximal rank. The associated a-diagram is just η'.

(β) and (β'): *We attach a to the row b left or right.* In this case the ab-diagram δ represents also a pair in which one of the two maps has maximal rank. The associated a-diagram is a one step degeneration of η'.

(γ): *We create a new row consisting of the remaining a.* In this case the associated a-diagram is a one step degeneration of η' but neither map in the nilpotent pair has maximal rank. On the other hand for this ab-diagram we have, in the notations of 5.3 $a_1 = b_1 = 1$ and hence $\Delta = 1$. For the corresponding nilpotent pair orbit we have thus the dimension

$$\dim X_\delta = \tfrac{1}{2}(\dim C_{\sigma'} + \dim C_\sigma) + \dim U \cdot \dim V - 1.$$

Summing up all this analysis we see by an easy induction, that we have proved:

If Z_λ projects to C_σ, σ a one step degeneration of η, then either

$$Z_\lambda \subseteq Z^0$$

or

$$\dim Z_\lambda \leqq \sum_{i=1}^{t-1} n_i n_{i+1} + \tfrac{1}{2}\dim C_\sigma - 1 \leqq \dim Z^0 - 2.$$

This completes the proof of 5.2.

5.7. Remark. The only place in which we have used characteristic zero, apart from the implication normal $\Rightarrow$ Cohen Macaulay, was in the proof of proposition 3.4, where we used the following fact: If V is a affine variety on which a reductive group G acts and W a closed subvariety invariant under G, then the induced map $W/G \to V/G$ is a *closed immersion* (cf. 1.4a)). In characteristic $p > 0$ one can only say that this map is finite and injective, i.e. W/G is purely inseparable over its image (cf. [21] § 4).

6. An Application to Tensor Representation

6.1. We present here the application due to Th. Vust announced in the introduction. Let $A \in \operatorname{End}(U)$ be a matrix, G_A the centralizer of A in $GL(U)$. We consider the action of G_A on the tensor space $U^{\otimes m}$ and wish to compute $\operatorname{End}_{G_A}(U^{\otimes m})$. One knows that $\operatorname{End}_{GL(U)}(U^{\otimes m})$ is spanned by the symmetric group $\mathfrak{S}_m$ acting on $U^{\otimes m}$ in the obvious way (cf. [2], [18]). Now clearly the endomorphisms $A^{h_1} \otimes A^{h_2} \otimes \ldots \otimes A^{h_m} \in \operatorname{End}(U^{\otimes m})$ also commute with G_A and we have:

Theorem (Th. Vust). *The algebra* $\mathrm{End}_{G_A}(U^{\otimes m})$ *is spanned by the elements*

$$\sigma \cdot A^{h_1} \otimes \ldots \otimes A^{h_m}, \qquad \sigma \in \mathfrak{S}_m,\ h_1, \ldots, h_m \in \mathbb{N}.$$

The proof will require some lemmas (mostly well known).

6.2. Let V be an affine variety, G a reductive group acting on V, $W \subseteq V$ a G-stable closed subvariety, M a linear representation of G and φ: $W \to M$ a G-equivariant morphism.

Lemma 1. *There exists a G-equivariant morphism* Φ: $V \to M$ *extending* φ.

Proof. Let $K[U]$, $K[W]$ be the coordinate rings of V, W. A G-equivariant morphism φ from W to M is given by an element $u \in (K[W] \otimes M)^G$. To extend φ to V is equivalent to lift u to $(K[V] \otimes M)^G$, and this is a simple consequence of linear reductivity. qed.

6.3. Let V be as before, $G := GL(V)$. We take now W to be the closure $\overline{GA}$ of an orbit GA for an element $A \in V$. We assume:

i) $\dim(\overline{GA} \setminus GA) \leq \dim GA - 2$,

ii) $\overline{GA}$ is a *normal* variety.

Let G_A denote the stabilizer of A in G and M be again a linear representation of G.

Lemma 2. *If* $B \in M$ *is invariant under* G_A, *then under the condition* i) *and* ii) *there exists a G-equivariant morphism* Φ: $V \to M$ *such that* $\Phi(A) = B$.

Proof. First of all we construct a morphism φ: $GA \to M$ given by $gA \mapsto gB$; this is well defined since $B \in M^{G_A}$. The two hypotheses i) and ii) on $\overline{GA}$ imply that φ extends (uniquely) to a G-equivariant map φ': $\overline{GA} \to M$. Finally taking $W = \overline{GA}$ and applying lemma 1 we have the required conclusion. qed.

6.4. We now want to apply these lemmas to the following set up: $M := \mathrm{End}(U^{\otimes m}) = \mathrm{End}(U)^{\otimes m}$, $A \in \mathrm{End}(U)$ the given matrix, $B \in \mathrm{End}_{G_A}(U^{\otimes m})$, $V := \mathrm{End}(U)$. To finish our proof it only remains to explicit the set of G-equivariant maps Φ: $\mathrm{End}(U) \to \mathrm{End}(U)^{\otimes m}$. This set can be easily computed (cf. [15]). We need two lemmas for which we refer to the literature.

Lemma 3 ([10] Lemma 4.9, [15] Theorem 1.2). *Let* $\sigma \in \mathfrak{S}_m$ *be decomposed into cycles:* $\sigma = (i_1\, i_2 \ldots i_k)\ (j_1 j_2 \ldots j_e) \ldots (t_1\, t_2 \ldots t_s)$ (including cycles of length one), *and* $Y = X_1 \otimes X_2 \otimes \ldots \otimes X_m \in \mathrm{End}(U)^{\otimes m}$. *Then*

$$\mathrm{Tr}(\sigma \cdot Y) = \mathrm{Tr}(X_{i_1} X_{i_2} \ldots X_{i_k}) \cdot \mathrm{Tr}(X_{j_1} X_{j_2} \ldots X_{j_e}) \ldots \mathrm{Tr}(X_{t_1} X_{t_2} \ldots X_{t_s}).$$

Lemma 4 ([16] Theorem 1, [15] Theorem 1.3). *The ring of invariants of the space of m-tuples of matrices* $(X_1, X_2, \ldots, X_m)$ *under simultaneous conjugation under* $GL(U)$ *is generated by the invariants*

$$\mathrm{Tr}(X_{v_1} X_{v_2} \ldots X_{v_k}), \qquad k \in \mathbb{N},\ v_1, \ldots, v_k \in \{1, 2, \ldots, m\}.$$

6.5. Let us now look at the space L of G-equivariant maps Φ: $\mathrm{End}(U) \to \mathrm{End}(U)^{\otimes m}$. Clearly L is a module over the *ring* R *of invariants of* $\mathit{End}(U)$.

Proposition. *L is spanned, as an R-module, by the maps of type:*

$$X \mapsto \sigma \cdot X^{h_1} \otimes X^{h_2} \otimes \ldots \otimes X^{h_m}, \qquad \sigma \in \mathfrak{S}_m,\ h_i \in \mathbb{N}.$$

Proof. Let Φ: $\mathrm{End}(U) \to \mathrm{End}(U)^{\otimes m}$ be a G-equivariant map. We introduce m new variables $Y_1, Y_2, \ldots, Y_m$ in $\mathrm{End}(U)$ and construct a function Ψ on $\mathrm{End}(U)^{m+1}$ by setting

$$\Psi(X, Y_1, Y_2, \ldots, Y_m) := \mathrm{Tr}(\Phi(X) \cdot Y_1 \otimes Y_2 \otimes \ldots \otimes Y_m).$$

By the non degeneracy of the trace form the mapping $\Phi \mapsto \Psi$ is an injection from L to the space of invariants of $X, Y_1, Y_2, \ldots, Y_m$ which are linear in $Y_1, Y_2, \ldots, Y_m$. Now by lemma 4 such invariants are of type

$$\sum t(X) \cdot \mathrm{Tr}(X^{h_1} Y_{i_1} X^{h_2} Y_{i_2} \ldots X^{h_k} Y_{i_k}) \cdot \mathrm{Tr}(X^{p_1} Y_{j_1} X^{p_2} Y_{j_2} \ldots) \ldots$$
$$\ldots \mathrm{Tr}(X^{s_1} Y_{t_1} X^{s_2} Y_{t_2} \ldots) \qquad (t(X) \in R).$$

The previous lemma 3 shows then that any such invariant is of type $\mathrm{Tr}(\Phi(X) \cdot Y_1 \otimes Y_2 \otimes \ldots \otimes Y_m)$, where $\Phi(X)$ is a linear combination with coefficients in R of the special maps $X \mapsto \sigma \cdot X^{h_1} \otimes X^{h_2} \otimes \ldots \otimes X^{h_m}$. The previous remark about the injectivity of $\Phi \mapsto \Psi$ completes the proof. qed.

6.6. We now sum all our work and prove the main theorem 6.1:

Let $B \in \mathrm{End}_{G_A}(U^{\otimes m}) = (\mathrm{End}(U)^{\otimes m})^{G_A}$; we have seen that there exists a G-equivariant map Φ:$\mathrm{End}(U) \to \mathrm{End}(U^{\otimes m})$ such that $\Phi(A) = B$. By the previous proposition 6.5 we know all equivariant maps. The very formula given by the proposition implies immediately the theorem. qed.

References

1. Borho, W., Kraft, H.: Über Bahnen und deren Deformationen bei linearen Aktionen reduktiver Gruppen. Comment. Math. Helv. **54**, 61–104 (1979)
2. De Concini, C., Procesi, C.: A characteristic free approach to invariant theory. Adv. Math. **21**, 330–354 (1976)
3. Donovan, P., Freislich, M.R.: The representation theory of finite graphs and associated algebras. Carleton Lecture Notes **5** (1973)
4. Elkik, R.: Désingularisation des adhérences d'orbites polarisables et des nappes dans les algèbres de Lie réductives. Preprint.
5. Gabriel, P.: Représentations indécomposables. Seminaire Bourbaki exp. **444**, Springer Lecture Notes **431** (1975)
6. Grothendieck, A., Dieudonné, J.: EGA 0–IV. Publ. Math. de l'I.H.E.S. **11**, **20**, **24**, **32**; Paris (1961–1967)
7. Hesselink, W.: Singularities in the nilpotent scheme of a classical group. Trans. Am. Math. Soc. **222**, 1–32 (1976)
8. Hesselink, W.: Closure of orbits in a Lie algebra. Comment. Math. Helv. **54**, 105–110 (1979)
9. Kempf, G., Knudsen, F., Mumford, D., Saint-Donat, B.: Toroidal Embeddings I. Springer Lecture Notes **339** (1973)
10. Kostant, B.: A theorem of Frobenius, a theorem of Amitsur-Levitzki and cohomology theory. J. Math. Mech. **7**, 237–264 (1958)
11. Kostant, B.: Lie group representations on polynomial rings. Am. J. Math. **86**, 327–402 (1963)

12. Kostant, B., Rallis, S.: Orbits and representations associated with symmetric spaces. Am. J. Math. **93**, 753–809 (1971)
13. Kraft, H.: Parametrisierung von Konjugationsklassen in $\mathfrak{sl}_n$. Math. Ann. **234**, 209–220 (1978)
14. Nazarova, L.A.: Representations of quivers of infinite type. Akad. Nauk. SSSR **37**, 752–791 (1973)
15. Procesi, C.: The invariant theory of $n \times n$ matrices. Adv. Math. **19**, 306–381 (1976)
16. Sibirskii, K.S.: On unitary and orthogonal matrix invariants. Dokl. Akad. Nauk SSSR **172** n° 1 (1967)
17. Vinberg, E.B.: The Weyl group of a graded Lie algebra. Izv. Akad. Nauk SSSR **40**, n° 3 (1976)
18. Weyl, H.: Classical groups. Princeton Math. Series **1** (1946)
19. Loupias, M.: Représentations indécomposables de dimension finie des algèbres de Lie. Manuscripta math. **6**, 365–379 (1972)
20. Mumford, D.: Geometric Invariant Theory. Erg. der Math. **34**, Berlin-Heidelberg-New York: Springer Verlag 1970
21. Demazure, M.: Démonstration de la conjecture de Mumford (d'après W. Haboush). Sém. Bourbaki 74/75, exp. **462.** Berlin-Heidelberg-New York: Springer Verlag, Lecture Notes **514** (1976)
22. Vust, Th.: Sur la théorie des invariants des groupes classiques. Ann. Inst. Fourier **26**, 1–31 (1976)
23. Procesi, C., Kraft, H.: Classi coniugati in $GL(n, \mathbb{C})$. Rend. Sem. mat. Univ. Roma (1979)

Received December 19, 1978/April 4, 1979

Inventiones math. 53, 249–253 (1979)

Inventiones mathematicae

Über eine Spezialschar von Modulformen zweiten Grades (II)

Hans Maaß

Hirtenaue 50, D-6900 Heidelberg, Bundesrepublik Deutschland

Im linearen Raum (Γ_2, k) der Modulformen zweiten Grades zur Siegelschen Modulgruppe Γ_2 und zum Gewicht $k \equiv 0 \pmod 2$ verdient eine Teilschar $\mathfrak{S}_k$ von Spitzenformen aus zahlentheoretischen Gründen ein besonderes Interesse. Eine Form

$$\chi(Z) = \sum_{T>0} a(T)\, e^{2\pi i \sigma(TZ)} \qquad (\sigma = \mathrm{Spur}) \tag{1}$$

dieser Schar ist gekennzeichnet durch die Koeffizientenrelationen

$$a(T) = \sum_{d/n,m,t} d^{k-1} a\begin{pmatrix} 1 & \frac{t}{2d} \\ \frac{t}{2d} & \frac{nm}{d^2} \end{pmatrix}, \qquad T = \begin{pmatrix} n & \frac{1}{2}t \\ \frac{1}{2}t & m \end{pmatrix}. \tag{2}$$

Dabei durchlaufe T alle halbganzen positiven Matrizen. Da der Koeffizient $a(T)$ nur von der Äquivalenzklasse von T abhängt: $a(T[U]) = a(T)$ für unimodulare Matrizen U, so besagen die Relationen (2) insbesondere, daß $a(T)$ nur von $e = \mathrm{g.g.T.}(m, n, t)$ und $|T|$ abhängt. Überdies ist die Form χ durch ihre Koeffizienten $a(T)$ zu primitiven $T (e=1)$ bereits eindeutig bestimmt. Auf Grund umfangreicher Koeffiziententabellen wurden die Relationen (2) zuerst von Resnikoff und Saldaña [4] für die Eisensteinreihe φ_k und Igusas Spitzenformen χ_{10} und χ_{12} als Vermutung formuliert. Die weiter reichende Vermutung Kurokawas [2], daß $\dim \mathfrak{S}_k = \left[\frac{k-4}{6}\right]$ für $k \geqq 4$ ist, wurde durch Koeffiziententabellen für Spitzenformen bis zum Gewicht $k=20$ motiviert. Gezeigt werden konnte bisher nur, daß die Koeffizienten der Eisensteinreihe den Relationen (2) genügen. Der von mir angegebene Beweis – er beruht auf dreifach iterierter Induktion – ist vergleichsweise kompliziert. Ein einfacherer Beweis bietet sich auch für φ_k auf der Grundlage der Überlegungen in [3] sowie der vorliegenden Note an.

0020-9910/79/0053/0249/$01.00

Unlängst gelang mir der Nachweis der Ungleichung $\dim \mathfrak{S}_k \leqq \left[\frac{k-4}{6}\right]$ für $k \geqq 4$. Offen blieb die Frage, ob es überhaupt Formen $\chi \neq 0$ in $\mathfrak{S}_k$ gibt. Mit den in [3] dargelegten Ansätzen läßt sich jedoch, wie hier gezeigt werden soll, nicht nur die Existenzfrage konstruktiv beantworten, sondern sogar

$$\dim \mathfrak{S}_k = \left[\frac{k-4}{6}\right] \quad \text{für } k \geqq 4 \tag{3}$$

beweisen. Damit werden die formulierten Vermutungen von Resnikoff, Saldaña und Kurokawa in vollem Umfang bestätigt. Insbesondere liegen Igusas Spitzenformen χ_{10} und χ_{12} in der Spezialschar $\mathfrak{S}_{10}$ bzw. $\mathfrak{S}_{12}$.

Die Konstruktion der Formen $\chi \in \mathfrak{S}_k$ stützt sich auf die Entwicklung von χ nach Jacobischen Formen:

$$\chi(Z) = \sum_{m=1}^{\infty} \Theta_m(z_1, z)\, e^{2\pi i m z_2}, \qquad Z = \begin{pmatrix} z_1 & z \\ z & z_2 \end{pmatrix}. \tag{4}$$

Tatsächlich ist χ bereits durch Θ_1 eindeutig bestimmt; in [3] wurde nämlich gezeigt, daß notwendig

$$\Theta_m(z_1, z) = \Theta_1 | T(m)(z_1, \sqrt{m}\, z) \tag{5}$$

ist, wenn die Wirkung des Heckeschen Operators $T(m)$ durch

$$\Theta_1 | T(m)(z_1, z) = m^{k-1} \sum_{\nu=1}^{r} \Theta_1 | S_\nu, \tag{6}$$

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| a, b, c, d \in \mathbb{Z}, ad - bc = m \right\} = \bigcup_{\nu=1}^{r} \Gamma_1 S_\nu \tag{7}$$

definiert wird, wobei Γ_1 die Modulgruppe ersten Grades bezeichnet und allgemein

$$\Theta | S(z_1, z) = \Theta \left(S(z_1), \frac{\sqrt{m}\, z}{c z_1 + d} \right) e^{-2\pi i \frac{c z^2}{c z_1 + d}} (c, z_1 + d)^{-k} \tag{8}$$

für reelle Substitutionen $S = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ mit der Determinante $m > 0$ gesetzt wird. Durch Rechnung bestätigt man $(\Theta | M) | S = \Theta | (MS)$. Die Unabhängigkeit der Operation (6) von der Auswahl des Repräsentantensystems (S_ν) ist gewährleistet, wenn

$$\Theta_1 | M = \Theta_1 \quad \text{für } M \in \Gamma_1 \tag{9}$$

gilt. Diese Transformationsinvarianz wiederum wird gemäß [1] durch Wahl von Θ_1 garantiert:

$$\Theta_1(z_1, z) = \sum_{h=0}^{1} c_h(z_1) \sum_{n=-\infty}^{\infty} e^{2\pi i \left\{ z_1 \left(n + \frac{h}{2}\right)^2 + 2\left(n + \frac{h}{2}\right) z \right\}}. \tag{10}$$

Dabei sind c_0 und c_1 ganze Formen zur Hauptkonkruenzgruppe Γ_1 [4], zum Gewicht $k-\frac{1}{2}$ und zu gewissen Multiplikatorsystemen. Hinsichtlich einer genauen Beschreibung sei auf [3] verwiesen. Die Fourierentwicklung von c_h ist jedenfalls vom Typus

$$c_h(z_1)=\sum_{n-\frac{h}{4}\geqq 0}\gamma\left(n-\frac{h}{4}\right)e^{2\pi i\left(n-\frac{h}{4}\right)z_1}\qquad (h=0,1;n\in\mathbb{Z}),\tag{11}$$

und die Fourierkoeffizienten $\gamma\left(n-\frac{h}{4}\right)$ wachsen für $n\to\infty$ höchstens wie eine feste Potenz von $n-\frac{h}{4}$. Die Konvergenz aller gebildeten unendlichen Reihen ist daher leicht einzusehen. Die Jacobische Form Θ_1 kann nur dann zu einer Spitzenform χ führen, wenn $\Theta_1(z_1,z)\to 0$ für $\operatorname{Im} z_1\to\infty$ gilt. Gleichwertig damit ist $\gamma(0)=0$. Der lineare Raum der so ausgezeichneten Jacobischen Formen Θ_1 hat nach [3] die Dimension $\left[\frac{k-4}{6}\right]$; denn c_1 ist, wie in [3] gezeigt wurde, durch c_0 eindeutig bestimmt. Wir brauchen also nur noch zu zeigen, daß die mit einer ausgezeichneten Form Θ_1 gebildete Reihe (4) in $\mathfrak{S}_k$ liegt.

Gemäß Ansatz gestattet Θ_1 eine Fourierentwicklung der Art

$$\Theta_1(z_1,z)=\sum_{n=1}^{\infty}\sum_{t=-\infty}^{\infty}\alpha(n,t)\,e^{2\pi i(nz_1+tz)},\tag{12}$$

wobei $\alpha(n,t)=0$ im Falle $t^2\geqq 4n$ ist. Um die Wirkung von $T(m)$ auf Θ_1 zu ermitteln, wählen wir das spezielle Repräsentantensystem $S=\begin{pmatrix}a & b\\ 0 & d\end{pmatrix}$ mit $ad=m$, $d>0$, $b \bmod d$. Es ergibt sich

$$\Theta_1|T(m)(z_1,z)=m^{k-1}\sum_S\sum_{n=1}^{\infty}\sum_{t=-\infty}^{\infty}\alpha(n,t)\,e^{2\pi i\left(n\frac{az_1+b}{d}+t\frac{\sqrt{m}\,z}{d}\right)}d^{-k},$$

also

$$\Theta_m(z_1,z)=\sum_{n=1}^{\infty}\sum_{t=-\infty}^{\infty}\{\sum_{a/m}a^{k-1}\alpha(nd,t)\}\,e^{2\pi i(naz_1+taz)}.$$

Ersetzt man hierin na, ta durch n, t, so nimmt Θ_m die Gestalt

$$\Theta_m(z_1,z)=\sum_{n=1}^{\infty}\sum_{t=-\infty}^{\infty}\left\{\sum_{a/m,n,t}a^{k-1}\alpha\left(\frac{mn}{a^2},\frac{t}{a}\right)\right\}e^{2\pi i(nz_1+tz)}\tag{13}$$

an. Auf Grund des Ansatzes (4) erweist sich $\chi(Z)$ nunmehr als eine in z_1, z_2 symmetrische Funktion. Überdies ist

$$\chi(Z+B)=\chi(Z)\qquad\text{für ganze symmetrische } B.\tag{14}$$

Mit Hilfe einer gegebenen Modulsubstitution $M=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\in\Gamma_1$ bilden wir die Modulsubstitutionen zweiten Grades

$$M_1=\begin{pmatrix}\alpha&0&\beta&0\\0&1&0&0\\\gamma&0&\delta&0\\0&0&0&1\end{pmatrix},\quad M_2=\begin{pmatrix}1&0&0&0\\0&\alpha&0&\beta\\0&0&1&0\\0&\gamma&0&\delta\end{pmatrix}$$

und definieren in üblicher Weise

$$\Psi\,|\,M^*(Z)=\Psi((AZ+B)(CZ+D)^{-1})\,|CZ+D|^{-k}\qquad \text{für } M^*=\begin{pmatrix}A&B\\C&D\end{pmatrix}\in\Gamma_2.$$

Schließlich setzen wir noch

$$\Psi_m(Z)=\Theta_m(z_1,z)\,e^{2\pi i m z_2}, \tag{15}$$

so daß

$$\chi(Z)=\sum_{m=1}^{\infty}\Psi_m(Z) \tag{16}$$

wird. Wir zeigen die Invarianz von Ψ_m bezüglich M_1:

$$\begin{aligned}
&\Psi_m|M_1(Z)\\
&=\Theta_m\left(M(z_1),\frac{z}{\gamma z_1+\delta}\right)e^{2\pi i m\left(z_2-\frac{\gamma z^2}{\gamma z_1+\delta}\right)}(\gamma z_1+\delta)^{-k}\\
&=\Theta_1|T(m)\left(M(z_1),\frac{\sqrt{m}\,z}{\gamma z_1+\delta}\right)e^{-2\pi i\frac{\gamma(\sqrt{m}\,z)^2}{\gamma z_1+\delta}}(\gamma z_1+\delta)^{-k}\,e^{2\pi i m z_2}\\
&=m^{k-1}\sum_S\Theta_1\Big|S\left(M(z_1),\frac{\sqrt{m}\,z}{\gamma z_1+\delta}\right)e^{-2\pi i\frac{\gamma(\sqrt{m}\,z)^2}{\gamma z_1+\delta}}(\gamma z_1+\delta)^{-k}\,e^{2\pi i m z_2}\\
&=m^{k-1}\sum_S(\Theta_1|S)|M(z_1,\sqrt{m}\,z)\,e^{2\pi i m z_2}\\
&=m^{k-1}\sum_S\Theta_1|(SM)(z_1,\sqrt{m}\,z)\,e^{2\pi i m z_2}\\
&=\Theta_1|T(m)(z_1,\sqrt{m}\,z)\,e^{2\pi i m z_2}=\Psi_m(Z).
\end{aligned}\tag{17}$$

Hier ist zu beachten, daß mit S auch SM ein Repräsentantensystem der in (7) angegebenen Art durchläuft. Zufolge (17) ist $\chi|M_1=\chi$ und wegen der Symmetrie von $\chi(Z)$ in z_1, z_2 auch $\chi|M_2=\chi$, also $\chi|M_1M_2=\chi$. Speziell für $M=\begin{pmatrix}0&1\\-1&0\end{pmatrix}$ wird $M_1M_2=\begin{pmatrix}0&E\\-E&0\end{pmatrix}$, wobei E die zweireihige Einheitsmatrix bezeichnet. Da Γ_2 aus den Modulsubstitutionen

$$\begin{pmatrix}E&B\\0&E\end{pmatrix}\ (B \text{ ganz, symmetrisch})\quad\text{und}\quad\begin{pmatrix}0&E\\-E&0\end{pmatrix}$$

erzeugt werden kann, erweist sich χ als eine Form zur Gruppe Γ_2 und zum Gewicht k. χ ist Spitzenform, da zufolge (13)

$$\chi(Z)=\sum_T a(T)\, e^{2\pi i \sigma(TZ)} \tag{18}$$

mit

$$a\begin{pmatrix} n & \frac{1}{2}t \\ \frac{1}{2}t & m \end{pmatrix} = \sum_{a/m,n,t} a^{k-1}\,\alpha\left(\frac{mn}{a^2}, \frac{t}{a}\right) \tag{19}$$

gilt, so daß $a(T)=0$ im Falle $T \not> 0$ wird. Insbesondere ist

$$a\begin{pmatrix} 1 & \frac{1}{2}t \\ \frac{1}{2}t & m \end{pmatrix} = \alpha(m,t).$$

Die Relationen (2) sind daher erfüllt; d.h. χ liegt in $\mathfrak{S}_k$, q.e.d.

Im folgenden sei $(\Gamma_n, k)_0$ für $n=1,2$ die Schar aller Spitzenformen in (Γ_n, k) und φ_k die normierte Eisensteinreihe in (Γ_2, k). Die in $(\Gamma_2, k)_0$ gelegenen Potenzprodukte $\chi_{10}^a \chi_{12}^b \varphi_4^c \varphi_6^d$ mit $a+b=1$ bzw. $a+b \geqq 2$ erzeugen einen Teilraum $\mathfrak{T}'_k$ bzw. $\mathfrak{T}''_k$. Wie schon in [2] ausgeführt wurde, ist $\dim \mathfrak{T}'_k = \left[\frac{k-4}{6}\right] = \dim(\Gamma_1, 2k-2)_0$ und $\mathfrak{S}_k \cap \mathfrak{T}''_k = \{0\}$. Ein Struktursatz von Igusa besagt ferner, daß $(\Gamma_2, k)_0 = \mathfrak{T}'_k \oplus \mathfrak{T}''_k$ ist. Nach einer Bemerkung des Referenten folgt hieraus bereits $\dim \mathfrak{S}_k \leqq \operatorname{codim} \mathfrak{T}''_k = \dim \mathfrak{T}'_k$, also das Hauptergebnis von [3]. Darüber hinaus wurde in [3] bewiesen, daß der lineare Raum der Jacobischen Formen Θ_1 mit der Transformationsinvarianz $\Theta_1 | M = \Theta_1$ für $M \in \Gamma_1$ die Dimension $\left[\frac{k+2}{6}\right]$ hat. Erst auf Grund dieser Aussage ist es möglich, wie in der vorliegenden Note ausgeführt ist, $\dim \mathfrak{S}_k = \left[\frac{k-4}{6}\right]$ zu beweisen. Offenbar ist $(\Gamma_2, k)_0 = \mathfrak{S}_k \oplus \mathfrak{T}''_k$; d.h. es gibt in $\mathfrak{S}_k$ eine Basis, bestehend aus Formen der Art

$$\chi_{10}^a \chi_{12}^b \varphi_4^c \varphi_6^d + \psi_{a,b,c,d} \qquad \text{mit } a+b=1 \text{ und } \psi_{a,b,c,d} \in \mathfrak{T}''_k.$$

Der Tatsache, daß $\dim \mathfrak{S}_k = \dim(\Gamma_1, 2k-2)_0$ ist, dürfte einer Vermutung Kurokawas zufolge [2] eine tiefliegende Bedeutung zukommen.

Literatur

1. Eichler, M.: Über die Anzahl der linear unabhängigen Siegelschen Modulformen von gegebenem Gewicht. Math. Ann. **213**, 281–291 (1975)
2. Kurokawa, N.: Examples of eigenvalues of Hecke operators on Siegel cusp forms of degree two. Inv. Math. **49**, 149–165 (1978)
3. Maaß, H.: Über eine Spezialschar von Modulformen zweiten Grades. Inv. Math. **52**, 95–104 (1979)
4. Resnikoff, H.L., Saldaña, R.L.: Some properties of Fourier coefficients of Eisenstein series of degree two. J. f. d. reine u. angew. Math. **265**, 90–109 (1974)

Eingegangen am 24. April 1979

Inventiones math. 53, 255–265 (1979)

Inventiones mathematicae

Über eine Spezialschar von Modulformen zweiten Grades (III)

Hans Maaß

Hirtenaue 50, D-6900 Heidelberg, Bundesrepublik Deutschland

Es bezeichne (Γ_2, k) den linearen Raum der Modulformen zur Siegelschen Modulgruppe zweiten Grades Γ_2 und zum Gewicht $k \equiv 0 \pmod 2$, analog (Γ, k) den linearen Raum der elliptischen Modulformen zum Gewicht k und $(\Gamma, k)_0$ den Teilraum der Spitzenformen in (Γ, k). Unter $\mathfrak{S}_k$ werde die in [6, 7] behandelte Spezialschar der Spitzenformen

$$\chi(Z) = \sum_{N>0} a(N)\, e^{2\pi i \sigma(NZ)} \in (\Gamma_2, k) \qquad (\sigma = \mathrm{Spur}) \tag{1}$$

verstanden, für welche die Koeffizientenrelationen

$$a\begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix} = \sum_{g/a,b,c} g^{k-1}\, a\begin{pmatrix} 1 & \dfrac{b}{2g} \\ \dfrac{b}{2g} & \dfrac{ac}{g^2} \end{pmatrix} \tag{2}$$

charakteristisch sind. Eine gewisse Vertrautheit mit den in [6, 7] verwendeten Bezeichnungen ist unerläßlich, da ich mich auf diese Arbeiten beziehe. In [7] wurde gezeigt, daß die Formen χ der Spezialschar $\mathfrak{S}_k$ mit den Formen des Typus

$$\chi(Z) = \sum_{m=1}^{\infty} \Theta_1 | T(m)(z_1, \sqrt{m}\, z)\, e^{2\pi i m z_2}, \qquad Z = \begin{pmatrix} z_1 & z \\ z & z_2 \end{pmatrix} \tag{3}$$

identisch sind. Die Jacobische Form Θ_1 hat die Gestalt

$$\Theta_1(z_1, z) = c_0(z_1)\, \vartheta_0(z_1, z) + c_1(z_1)\, \vartheta_1(z_1, z) \tag{4}$$

mit

$$\vartheta_h(z_1, z) = \sum_{n=-\infty}^{\infty} e^{2\pi i \left\{\left(n+\frac{h}{2}\right)^2 z_1 + (2n+h) z\right\}} \qquad \text{für } h = 0, 1 \tag{5}$$

0020-9910/79/0053/0255/$02.20

und gewissen Spitzenformen

$$c_h(z_1)=\sum_{n=1}^{\infty} \gamma_h\left(n-\frac{h}{4}\right) e^{2\pi i\left(n-\frac{h}{4}\right)z_1} \in (\Gamma_0(4), k-\tfrac{1}{2}, v_h). \tag{6}$$

Da c_1, wie in [6] ausgeführt wurde, durch c_0 eindeutig bestimmt ist, so ergibt sich eine umkehrbar eindeutige Beziehung $\chi \leftrightarrow \Theta_1 \leftrightarrow c_0$.

In der vorliegenden Arbeit wird gezeigt, daß das Eulerprodukt $D(s,\chi)$ einer Eigenfunktion $\chi \in \mathfrak{S}_k$ der Heckeschen Operatoren $T(n)$ mit den Eigenwerten $\lambda(n)$ im Sinne der Theorie Andrianovs [1] die Gestalt

$$D(s,\chi)=\zeta(s-k+1)\,\zeta(s-k+2)\,D(s) \tag{7}$$

hat, wobei $\zeta(s)$ die Riemannsche Zetafunktion und

$$D(s)=\prod_p (1-\omega(p)\,p^{-s}+p^{2k-3-2s})^{-1}, \qquad \omega(p)=\lambda(p)-p^{k-1}-p^{k-2} \tag{8}$$

ist. Das Produkt ist hier über alle Primzahlen p zu erstrecken. Vermöge (7) erweist sich $D(s)$ als eine meromorphe Funktion von s. Sie genügt der Funktionalgleichung

$$\Psi(s)=(2\pi)^{-s}\,\Gamma(s)\,D(s)=-\Psi(2k-2-s), \tag{9}$$

die sich unmittelbar aus der von Andrianov [1] abgeleiteten Funktionalgleichung für $D(s,\chi)$:

$$\Psi(s,\chi)=(2\pi)^{-2s}\,\Gamma(s)\,\Gamma(s-k+2)\,D(s,\chi)=\Psi(2k-2-s,\chi) \tag{10}$$

und der entsprechenden für $\zeta(s)$ ergibt. Ist $D(s)$ eine ganze Funktion, so folgt aus den Sätzen [4, § 2] und [5, Satz 42], daß die der Reihe

$$D(s)=\sum_{n=1}^{\infty} a_n n^{-s} \text{ zugeordnete Funktion } f(z)=\sum_{n=1}^{\infty} a_n e^{2\pi i n z} \tag{11}$$

eine Spitzenform in $(\Gamma, 2k-2)_0$ und überdies Eigenfunktion aller Heckeschen Operatoren $T(n)$ zu den Eigenwerten $\omega(n)$ ist. $D(s)$ ist dann mit dem Eulerprodukt $D(s,f)$ der Form f identisch. Wir beweisen mit Hilfe der Theorie Shimuras [8], daß $D(s)$ eine ganze Funktion ist, falls $\chi(Z)$ einen Koeffizienten

$a(N)\neq 0$ zu primitivem N mit $|N|=t\,n^2$ besitzt, wobei t, n natürliche Zahlen sind, t quadratfrei und $t+1 \not\equiv 0 \pmod 4$ ist. (12)

Als Schlüssel für das Verständnis der analytischen Zusammenhänge erweisen sich die Relationen

$$\gamma_h\left(n-\frac{h}{4}\right)=a\begin{pmatrix}1 & \frac{1}{2}h\\ \frac{1}{2}h & n\end{pmatrix} \qquad \text{für } h=0,1 \quad \text{und} \quad n\geqq 1, \tag{13}$$

die man mit Hilfe eines einfachen Integrationsprozesses gewinnt. Sie besagen

$$\sum_{n=1}^{\infty} a\begin{pmatrix} 1 & \frac{1}{2}h \\ \frac{1}{2}h & n \end{pmatrix} e^{2\pi i\left(n-\frac{h}{4}\right)z} \in (\Gamma_0(4), k-\tfrac{1}{2}, v_h) \qquad \text{für } h=0,1. \tag{14}$$

Da $\mathfrak{S}_k$ für $k \leqq 18$ aus allen Spitzenformen in (Γ_2, k) besteht, so gibt es in diesen Fällen eine Basis $\chi_1, \chi_2, \dots, \chi_r$ von $\mathfrak{S}_k$, die aus Eigenfunktionen bezüglich der Heckeschen Operatoren besteht. Auf Grund der Ergebnisse von Kurokawa [2] ist dies zufolge $\dim \mathfrak{S}_{20} = 2$ auch noch für $k=20$ richtig. Schließlich zeigen die Rechnungen von Kurokawa für $k \leqq 20$, daß die Eulerprodukte $D(s, \chi_\nu)$ der Basisformen χ_ν paarweise verschieden sind; denn sie unterscheiden sich bereits im 2-Faktor, da die Eigenwerte $\lambda(2) = \lambda(2, \chi_\nu)$ paarweise verschieden sind. Die Forderung (12) ist in den genannten Fällen stets für die Einheitsmatrix $N = E$ erfüllt. Den Formen $\chi_1, \chi_2, \dots, \chi_r \in \mathfrak{S}_k$ entsprechen daher Spitzenformen $f_1, f_2, \dots, f_r \in (\Gamma, 2k-2)_0$ mit paarweise verschiedenen Eulerprodukten $D(s, f_\nu)$. Es besteht daher eine umkehrbar eindeutige Beziehung $\chi_\nu \leftrightarrow f_\nu$ $(\nu = 1, 2, \dots, r)$, die in den Relationen

$$D(s, \chi_\nu) = \zeta(s-k+1)\,\zeta(s-k+2)\,D(s, f_\nu) \qquad \text{für } \nu = 1, 2, \dots, r \tag{15}$$

ihren Ausdruck findet. Die „Conjecture 1" von Kurokawa [2] hat sich damit jedenfalls für die Gewichte $10 \leqq k \leqq 20$ als richtig erwiesen. Der Hinweis erscheint angezeigt, daß Hiroshi Saito mir diese Vermutung bereits am 16. 3. 1977 brieflich mitgeteilt hat. Es fehlte indessen der Bezug auf die Formenschar $\mathfrak{S}_k$, der sich hier als wesentlich herausgestellt hat.

Die folgenden Beweisansätze stützen sich vor allem auf Eigenschaften der Spitzenform c_0. Bringt man, wie Andrianov im Anschluß an die vorliegende Untersuchung während seines Heidelberger Seminars im Sommersemester 1979 vorschlug, mit c_0 gleichzeitig die Form c_1 ins Spiel, so läßt sich ohne zusätzliche Voraussetzung zeigen, daß die durch (8) erklärte Funktion ganz ist. Schließlich konnte Andrianov noch beweisen, daß die Schar $\mathfrak{S}_k$ gegenüber den Heckeschen Operatoren invariant ist. Er bereitet eine Publikation dieser Ergebnisse vor.

§ 1. Das Eulerprodukt einer Eigenfunktion $\chi \in \mathfrak{S}_k$

Wir vereinbaren folgende Bezeichnungen:

$$Z = X + iY, \qquad X = \begin{pmatrix} x_1 & x \\ x & x_2 \end{pmatrix}, \qquad Y = \begin{pmatrix} y_1 & y \\ y & y_2 \end{pmatrix}, \qquad dv = |Y|^{-\frac{3}{2}}\, dy_1\, dy_2\, dy$$

und setzen

$$\Gamma_2(s) := \int_{Y>0} e^{-\sigma(Y)} |Y|^s\, dv = \sqrt{\pi}\, \Gamma(s)\, \Gamma(s-\tfrac{1}{2}) \qquad (\sigma = \text{Spur}).$$

Zum Beweis der Relationen (13) berechnen wir das Integral

$$I_h(s) = \sum_{m=1}^{\infty} \int_{Y>0} \iiint_0^1 \chi(Z)\, e^{-2\pi i \sigma(N_h(m) \bar{Z})} |Y|^s\, dx_1\, dx_2\, dx\, dv \tag{16}$$

auf zwei verschiedene Weisen. Zur Abkürzung ist hier

$$N_h(m)=\begin{pmatrix}1 & \frac{1}{2}h\\ \frac{1}{2}h & m\end{pmatrix} \qquad \text{für } h=0,1 \quad \text{und} \quad m\in\mathbb{N}$$

gesetzt worden. Trägt man in (16) für χ die Fourierreihe (1) ein, so ergibt sich

$$\begin{aligned} I_h(s) &= \sum_{m=1}^{\infty} a\begin{pmatrix}1 & \frac{1}{2}h\\ \frac{1}{2}h & m\end{pmatrix} \int_{Y>0} e^{-4\pi\sigma(N_h(m)Y)}|Y|^s\,dv \\ &=(4\pi)^{-2s}\Gamma_2(s)\sum_{m=1}^{\infty} a\begin{pmatrix}1 & \frac{1}{2}h\\ \frac{1}{2}h & m\end{pmatrix}\left(m-\frac{h}{4}\right)^{-s} \qquad \text{für } h=0,1. \end{aligned} \tag{17}$$

Etwas länger gestaltet sich die Rechnung, wenn in (16) für χ die Entwicklung (3) eingetragen wird. Die Integration über x_2 führt auf

$$I_h(s)=\sum_{m=1}^{\infty}\int_{Y>0}\iint_0^1 \Theta_1|T(m)(z_1,\sqrt{m}\,z)\,e^{-2\pi i(\bar z_1+h\bar z)-4\pi m y_2}\,|Y|^s\,dx_1\,dx\,dv.$$

Um die Wirkung des Heckeschen Operators $T(m)$ auf Θ_1 zu ermitteln, wählen wir die Matrizen $S=\begin{pmatrix}a & b\\ 0 & d\end{pmatrix}$, $ad=m$, $b \bmod d$ als Repräsentantensystem. Die Reihendarstellung (5) ergibt dann

$$\begin{aligned} \Theta_1|T(m)(z_1,\sqrt{m}\,z) &= m^{k-1}\sum_S \Theta_1|S(z_1,\sqrt{m}\,z) \\ &= m^{k-1}\sum_S \Theta_1\left(\frac{a z_1+b}{d},\frac{m z}{d}\right)d^{-k} \\ &= m^{k-1}\sum_S\sum_{n=-\infty}^{\infty}\left\{c_0\left(\frac{a z_1+b}{d}\right)e^{2\pi i\left\{n^2\frac{a z_1+b}{d}+2naz\right\}}\right. \\ &\quad \left.+c_1\left(\frac{a z_1+b}{d}\right)e^{2\pi i\left\{(n+\frac{1}{2})^2\frac{a z_1+b}{d}+(2n+1)az\right\}}\right\}d^{-k}. \end{aligned}$$

Die Integration über x wird in den Fällen $h=0,1$ separat ausgeführt:

$$\begin{aligned} \int_0^1 \Theta_1|T(m)(z_1,\sqrt{m}\,z)\,dx &= m^{k-1}\sum_S c_0\left(\frac{a z_1+b}{d}\right)d^{-k} \\ &= m^{k-1}\sum_S\sum_{n=1}^{\infty}\gamma_0(n)\,e^{2\pi i\frac{a z_1+b}{d}n}\,d^{-k} \\ &= \sum_{d/m}\sum_{n=1}^{\infty} a^{k-1}\,\gamma_0(n d)\,e^{2\pi i n a z_1}, \end{aligned}$$

$$\int_0^1 \Theta_1 | T(m)(z_1, \sqrt{m}\, z)\, e^{-2\pi i \bar{z}}\, dx$$

$$= \frac{1}{m} \sum_{b \bmod m} c_1\left(\frac{z_1+b}{m}\right) e^{2\pi i \frac{z_1+b}{4m}} e^{-4\pi y}$$

$$= \frac{1}{m} \sum_{b \bmod m} \sum_{n=1}^{\infty} \gamma_1(n-\tfrac{1}{4})\, e^{2\pi i n \frac{z_1+b}{m}} e^{-4\pi y}$$

$$= \sum_{n=1}^{\infty} \gamma_1(n m - \tfrac{1}{4})\, e^{2\pi i n z_1} e^{-4\pi y}.$$

Einheitlich für $h=0,1$ folgt nun

$$\iint_0^1 \Theta_1 | T(m)(z_1, \sqrt{m}\, z)\, e^{-2\pi i(\bar{z}_1 + h\bar{z})}\, dx_1\, dx = \gamma_h\left(m - \frac{h}{4}\right) e^{-4\pi(y_1 + hy)}$$

und damit

$$I_h(s) = \sum_{m=1}^{\infty} \gamma_h\left(m - \frac{h}{4}\right) \int_{Y>0} e^{-4\pi(y_1 + hy + my_2)} |Y|^s\, dv$$

$$= (4\pi)^{-2s}\, \Gamma_2(s) \sum_{m=1}^{\infty} \gamma_h\left(m - \frac{h}{4}\right)\left(m - \frac{h}{4}\right)^{-s}. \tag{18}$$

Vergleich mit (17) ergibt in der Tat die Relationen (13).

Wir bestimmen die Wirkung des Heckeschen Operators $T(p)$, wobei p eine beliebige Primzahl bezeichnet, auf eine Spitzenform $\chi \in \mathfrak{S}_k$. Es seien $a(p, N)$ die Fourierkoeffizienten von $\chi | T(p)$. Bekanntlich (s. [1]) ist

$$a(p, N) = p^{2k-3} \sum_{\substack{ghd=p \\ U}} a\left(\frac{g}{hd} N\left[U\begin{pmatrix}1 & 0\\ 0 & d\end{pmatrix}\right]\right) g^{3-2k} d^{1-k}.$$

Summiert wird hier über alle multiplikativen Zerlegungen $g h d = p$ in natürliche Zahlen, und U durchläuft bei gegebenem d ein vollständiges System von unimodularen Matrizen, deren erste Spalten $\bmod d$ nicht assoziiert sind. Wir wählen

$$d=1:\ U=E \qquad \text{(Einheitsmatrix)},$$

$$d=p:\ U=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\ \nu & 1\end{pmatrix} \qquad (0 \leqq \nu < p).$$

Speziell für $N = \begin{pmatrix}1 & 0\\ 0 & m\end{pmatrix}$ ergibt sich

$$a\left(p, \begin{pmatrix}1 & 0\\ 0 & m\end{pmatrix}\right) = a\begin{pmatrix}p & 0\\ 0 & pm\end{pmatrix} + p^{k-2}\, a\begin{pmatrix}mp^{-1} & 0\\ 0 & p\end{pmatrix}$$

$$+ p^{k-2} \sum_{\nu=0}^{p-1} a\begin{pmatrix}(1+\nu^2 m)\,p^{-1} & \nu m\\ \nu m & pm\end{pmatrix},$$

wenn allgemein $a(N)=0$ für nicht halbganze und nicht positive N gesetzt wird. Entsprechend sei auch $\gamma_0(x)=\gamma_1(x-\frac{1}{4})=0$ für $x\notin\mathbb{N}$. Fortgesetzte Anwendung von (2) und (13) ergibt für $p>2$ die Beziehung

$$a\left(p,\begin{pmatrix}1 & 0\\ 0 & m\end{pmatrix}\right)=\gamma_0(p^2 m)+\left\{p^{k-1}+p^{k-2}+\left(\frac{-m}{p}\right)p^{k-2}\right\}\gamma_0(m)+p^{2k-3}\gamma_0\left(\frac{m}{p^2}\right). \tag{19}$$

Im Fall $p=2$ gilt,

$$a\left(2,\begin{pmatrix}1 & 0\\ 0 & m\end{pmatrix}\right)=a\begin{pmatrix}2 & 0\\ 0 & 2m\end{pmatrix}+2^{k-2}a\begin{pmatrix}\frac{m}{2} & 0\\ 0 & 2\end{pmatrix}+2^{k-2}a\begin{pmatrix}\frac{m+1}{2} & m\\ m & 2m\end{pmatrix}$$

$$=\gamma_0(4m)+(2^{k-1}+2^{k-2})\gamma_0(m)+2^{2k-3}\left(\gamma_0\left(\frac{m}{4}\right)+\gamma_1\left(\frac{m}{4}\right)\right). \tag{20}$$

Schließlich bestimmen wir noch

$$a\left(2,\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & m\end{pmatrix}\right)=a\begin{pmatrix}2 & 1\\ 1 & 2m\end{pmatrix}+2^{k-2}a\begin{pmatrix}\frac{m}{2} & \frac{1}{2}\\ \frac{1}{2} & 2\end{pmatrix}+2^{k-2}a\begin{pmatrix}\frac{m}{2}+1 & m+\frac{1}{2}\\ m+\frac{1}{2} & 2m\end{pmatrix}$$

$$=\gamma_0(4m-1)+2^{k-1}\gamma_1(m-\tfrac{1}{4})+\delta\left(\frac{m}{2}\right)2^{k-1}\gamma_1(m-\tfrac{1}{4}). \tag{21}$$

Dabei ist $\delta(x)=1$ für $x\in\mathbb{Z}$ und 0 sonst.

Fortan sei χ eine Eigenfunktion, also

$$\chi\,|\,T(p)=\lambda(p)\,\chi \qquad \text{für alle Primzahlen } p\geqq 2. \tag{22}$$

Mit $\omega(p)=\lambda(p)-p^{k-1}-p^{k-2}$ ergibt sich dann

$$\omega(p)\,\gamma_0(m)=\gamma_0(p^2 m)+\left(\frac{-m}{p}\right)p^{k-2}\gamma_0(m)+p^{2k-3}\gamma_0\left(\frac{m}{p^2}\right) \qquad \text{für } p>2, \tag{23}$$

$$\omega(2)\,\gamma_0(m)=\gamma_0(4m)+2^{2k-3}\left(\gamma_0\left(\frac{m}{4}\right)+\gamma_1\left(\frac{m}{4}\right)\right), \tag{24}$$

$$(\lambda(2)-2^{k-1})\,\gamma_1(m-\tfrac{1}{4})=\gamma_0(4m-1)+\delta\left(\frac{m}{2}\right)2^{k-1}\gamma_1(m-\tfrac{1}{4}). \tag{25}$$

Die Koeffizienten von c_1 könnten mit Hilfe von (25) unmittelbar durch die von c_0 ausgedrückt werden, wenn sicher wäre, daß $\lambda(2)\neq 2^k$, 2^{k-1} ist. Eine allgemeine Aussage darüber ist jedoch nicht möglich. Mit χ ist auch c_0 von Null verschieden. Es gibt daher natürliche Zahlen t, die außer 1 keinen quadratischen Teiler haben, so daß

$$D_t(s) = \sum_{n=1}^{\infty} \gamma_0(t\,n^2)\,n^{-s} \neq 0. \tag{26}$$

Mit Hilfe von (23) läßt sich in Analogie zu [8, Theorem 1.9] für $D_t(s)$ die Produktdarstellung

$$D_t(s) = H_t(2^{-s}) \prod_{p>2} \left(1 - \left(\frac{-t}{p}\right) p^{k-2-s}\right) (1-\omega(p)\,p^{-s} + p^{2k-3-2s})^{-1} \tag{27}$$

mit

$$H_t(x) = \sum_{h=0}^{\infty} \gamma_0(t\,2^{2h})\,x^h$$

gewinnen. Für diese Reihe liefert (24), hier in der speziellen Form

$$\omega(2)\,\gamma_0(t\,2^{2h}) = \gamma_0(t\,2^{2h+2}) + 2^{2k-3}\{\gamma_0(t\,2^{2h-2}) + \gamma_1(t\,2^{2h-2})\},$$

einen geschlossenen Ausdruck. Multipliziert man diese Relation mit x^{h+1} und summiert man über $h \geqq 0$, so ergibt sich

$$H_t(x)(1-\omega(2)\,x + 2^{2k-3}\,x^2) = \gamma_0(t) - 2^{2k-3}\,\gamma_1\left(\frac{t}{4}\right)x, \tag{28}$$

denn es ist $\gamma_1(n) = 0$ für $n \in \mathbb{N}$.

Wir setzen nun $-4t = dq^2$. Dabei bezeichne d die Diskriminante des quadratischen Zahlkörpers $\mathbb{Q}(\sqrt{-t})$. Offenbar stimmt $\left(\frac{-t}{p}\right)$ im Falle $p > 2$ mit dem Restsymbol $\left(\frac{d}{p}\right)$ überein. (27) und (28) ergeben daher

$$L\left(s-k+2, \left(\frac{d}{*}\right)\right) D_t(s) = D(s) \left(1 - \left(\frac{d}{2}\right) 2^{k-2-s}\right)^{-1} \left(\gamma_0(t) - 2^{2k-3-s}\,\gamma_1\left(\frac{t}{4}\right)\right) \tag{29}$$

mit

$$D(s) = \prod_{p \geqq 2} (1-\omega(p)\,p^{-s} + p^{2k-3-2s})^{-1}. \tag{30}$$

Für das Eulerprodukt $D(s,\chi)$ der Form χ hat Andrianov folgende Darstellung gegeben

$$L_{-4t}(s-k+2) \sum_{i=1}^{h} \sum_{m=1}^{\infty} \frac{a(m\,N_i)}{m^s} = \Phi_\chi(s)\,D(s,\chi). \tag{31}$$

Die hier auftretenden Größen haben gemäß [1, Theorem 2.4.1] folgende Bedeutung: $N_i = \begin{pmatrix} a_i & b_i \\ b_i & c_i \end{pmatrix}$, $i = 1, 2, \ldots, h$, bezeichnet ein vollständiges System von Repräsentanten der Klassen der im engeren Sinne äquivalenten positiv definiten primitiven Matrizen mit der Determinante t. Demgemäß ist b_i ganz und g.g.T. $(a_i, 2b_i, c_i)=1$. Anstelle des in den Andrianovschen Formeln auftretenden allgemeinen Gruppencharakters steht hier der Einheitscharakter. Er tritt daher als Argument nicht mehr in Erscheinung. Schließlich ist

$$L_{-4t}(s)=\prod_{\mathfrak{p}\nmid q}(1-N\mathfrak{p}^{-s})^{-1}=\prod_{p\nmid q}(1-p^{-s})^{-1}\left(1-\left(\frac{d}{p}\right)p^{-s}\right)^{-1}$$
$$=\zeta(s)\,L\left(s,\left(\frac{d}{*}\right)\right)\prod_{p/q}(1-p^{-s})\left(1-\left(\frac{d}{p}\right)p^{-s}\right) \tag{32}$$

und

$$\Phi_\chi(s)=\sum_{i=1}^{h}\left\{\prod_{p/q}\left(1-\frac{\pi(p)}{p^{s-k+2}}\right)\left(1-\frac{\Delta^-(p)}{p^{s-2k+3}}\right)a\right\}(N_i). \tag{33}$$

Die Operatoren $\pi(p)$ und $\Delta^-(p)$ sind durch [1, (2.1.14) und (2, 1, 17)] erklärt. In dem ersten Produkt in (32) durchläuft $\mathfrak{p}$ alle Primideale des Körpers $\mathbb{Q}(\sqrt{d})$, die q nicht teilen. $N\mathfrak{p}$ bezeichnet die Norm von $\mathfrak{p}$.

Auf Grund der Relationen (2) ergibt sich

$$\sum_{m=1}^{\infty} a(mN_i)\,m^{-s}=\zeta(s-k+1)\,D_t(s), \tag{34}$$

unabhängig von i; denn es ist

$$a(mN_i)=\sum_{d/m} d^{k-1}\,\gamma_0(tm^2d^{-2}),$$

also

$$\sum_{m=1}^{\infty} a(mN_i)\,m^{-s}=\sum_{m,n=1}^{\infty} n^{k-1-s}\,\gamma_0(tm^2)\,m^{-s}.$$

Aus (29), (31), (32), (34) folgt nun

$$h\,\zeta(s-k+1)\,\zeta(s-k+2)\,D(s)\left\{\gamma_0(t)-2^{2k-3-s}\gamma_1\left(\frac{t}{4}\right)\right\}\prod_{p/q}(1-p^{k-2-s})$$
$$=\Phi_\chi(s)\,D(s,\chi). \tag{35}$$

Die behauptete Darstellung

$$D(s,\chi)=\zeta(s-k+1)\,\zeta(s-k+2)\,D(s) \tag{36}$$

reduziert sich demnach auf den Nachweis von

$$\Phi_\chi(s)=h\left\{\gamma_0(t)-2^{2k-3-s}\gamma_1\left(\frac{t}{4}\right)\right\}\prod_{p/q}(1-p^{k-2-s}). \tag{37}$$

Wir weisen ausdrücklich darauf hin, daß zufolge (26) und (29) die Koeffizienten $\gamma_0(t)$ und $\gamma_1\left(\frac{t}{4}\right)$ nicht gleichzeitig verschwinden, so daß in (35) eine Kürzung durch $\Phi_\chi(s)$ möglich ist.

Ist $q=1$, also $d=-4t$, so folgt $-t\not\equiv 1 \pmod 4$, also $t+1\not\equiv 0 \pmod 4$, mithin $\gamma_1\left(\frac{t}{4}\right)=0$. Die rechte Seite von (37) reduziert sich also auf $h\gamma_0(t)$ in Übereinstimmung mit $\Phi_\chi(s)=\sum_{i=1}^{h} a(N_i)=h\gamma_0(t)$. Es bleibt der Fall $q=2$ zu untersuchen. Nun ist $d=-t\equiv 1 \pmod 4$. Gemäß (37) wird dann

$$\begin{aligned}\sum_{i=1}^{h}\{a(N_i)-2^{k-2-s}(\pi(2)\,a)(N_i)-2^{2k-3-s}(\varDelta^-(2)\,a)(N_i)\\ +2^{3k-5-2s}(\pi(2)\,\varDelta^-(2)\,a)(N_i)\}\\ =h\left\{\gamma_0(t)-2^{k-2-s}\gamma_0(t)-2^{2k-3-s}\gamma_1\left(\frac{t}{4}\right)+2^{3k-5-2s}\gamma_1\left(\frac{t}{4}\right)\right\}\end{aligned}\tag{38}$$

behauptet. Wir setzen für einen fest gewählten Index $N=N_i=\begin{pmatrix}a & b\\ b & c\end{pmatrix}$. Da N primitiv ist, ist jedenfalls $a(N)=\gamma_0(t)$ und $(\varDelta^-(2)\mathrm{a})(N)=a(\frac{1}{2}N)=0$. Wegen $t=ac-b^2\equiv -1 \pmod 4$ und g.g.T. $(a,2b,c)=1$ sind im folgenden drei Fälle zu erörtern:

1) $b\equiv 0 \pmod 2$, $a\equiv c\equiv 1 \pmod 2$, $a+c\equiv 0 \pmod 4$,
2) $b\equiv 1 \pmod 2$, $a\equiv 0 \pmod 4$, $c\equiv 1 \pmod 2$,
3) $b\equiv 1 \pmod 2$, $a\equiv 1 \pmod 2$, $c\equiv 0 \pmod 4$.

Wir bestimmen die auf der linken Seite von (38) noch auftretenden Größen:

$$(\pi(2)\,a)(N)=a\begin{pmatrix}\frac{1}{2}a & b\\ b & 2c\end{pmatrix}+a\begin{pmatrix}\frac{1}{2}c & b\\ b & 2a\end{pmatrix}+a\begin{pmatrix}\frac{1}{2}(a+c)+b & b+c\\ b+c & 2c\end{pmatrix}.$$

In allen drei Fällen ist genau eine der hier auftretenden Matrizen halbganz; diese ist dann jeweils das Doppelte einer primitiven Matrix. Demnach ist

$$(\pi(2)\,a)(N)=\gamma_0(t)+2^{k-1}\gamma_1\left(\frac{t}{4}\right).$$

Zugleich ergibt sich nun aber auch

$$\begin{aligned}&(\pi(2)\,\varDelta^-(2)\,a)(N)\\ &=a\begin{pmatrix}\frac{1}{4}a & \frac{1}{2}b\\ \frac{1}{2}b & c\end{pmatrix}+a\begin{pmatrix}\frac{1}{4}c & \frac{1}{2}b\\ \frac{1}{2}b & a\end{pmatrix}+a\begin{pmatrix}\frac{1}{4}(a+2b+c) & \frac{1}{2}(b+c)\\ \frac{1}{2}(b+c) & c\end{pmatrix}=\gamma_1\left(\frac{t}{4}\right).\end{aligned}$$

Damit ist (38) verifiziert.

§ 2. Der Bezug zu den Eigenfunktionen $f\in(\Gamma, 2k-2)_0$

Die bisherigen Überlegungen wurden unabhängig von [8] ausgeführt. Um nun zu beweisen, daß das Produkt (30) unter der Voraussetzung (12) Eulerprodukt einer Eigenfunktion $f\in(\Gamma,2k-2)_0$ ist, müssen wir uns wesentlich auf Shimuras Ergebnisse stützen. Bezeichnet $(\Gamma_0(4),\ k-\frac{1}{2},v_h)_0$ den Raum der Spitzenformen in $(\Gamma_0(4),\ k-\frac{1}{2},v_h)$, so ist zu zeigen, daß $(\Gamma_0(4),\ k-\frac{1}{2},v_0)_0$ mit Shimuras Raum $S_{2k-1}(4,\varepsilon)$ identisch ist, wobei ε den Einheitscharakter mod 4 bezeichnet, und ferner, daß c_0 diesem Raum angehört. Daß c_0 eine Spitzenform darstellt, ist evident auf Grund der Entwicklungen (6), da $\begin{pmatrix} c_0 \\ c_1 \end{pmatrix}$ eine vektorielle Modulform zur Modulgruppe Γ ist, wie in [6] ausgeführt wurde. Es bleibt zu zeigen, daß

$$c_0(M(z))=j(M,z)^{2k-1}\,c_0(z) \qquad \text{für } M\in\Gamma_0(4) \tag{39}$$

gilt, wenn j durch

$$j(M,z)=\vartheta_0(M(z))/\vartheta_0(z) \text{ für } M\in\Gamma_0(4)$$

erklärt wird und allgemein

$$\vartheta_h(z)=\sum_{n=-\infty}^{\infty} e^{2\pi i\left(n+\frac{h}{2}\right)^2 z} \qquad \text{für } h=0,1 \tag{40}$$

gesetzt wird.

Mit gewissen Multiplikatorsystemen w_h ist

$$\vartheta_h\in(\Gamma_0(4),\tfrac{1}{2},w_h), \qquad w_h^4=1 \quad \text{für } h=0,1. \tag{41}$$

Gemäß [5, Satz 3] sowie den Ausführungen in [6, S. 102] bestehen Relationen

$$c_0\,\vartheta_0+c_1\,\vartheta_1=u, \qquad c_0\,\vartheta_0'+c_1\,\vartheta_1'=\frac{1}{2k}u'+\pi\,i\,v \tag{42}$$

mit gewissen Formen $u\in(\Gamma,k)$, $v\in(\Gamma,k+2)$. Die Form $\pi\,i\,\eta^6=\vartheta_0\,\vartheta_1'-\vartheta_1\,\vartheta_0'$ verschwindet nicht und gehört dem Raum $(\Gamma_0(4),\ 3,\ w_0\,w_1)$ an. Die Auflösung von (42) nach c_0 ergibt

$$c_0=\frac{1}{\pi\,i\,k\,\eta^6}(k\,\vartheta_1'\,u-\tfrac{1}{2}\vartheta_1\,u')-\frac{\vartheta_1\,v}{\eta^6}. \tag{43}$$

Diese Darstellung zeigt, daß c_0 und ϑ_0^{2k-1} ein und demselben Raum $(\Gamma_0(4),\ k-\frac{1}{2},\ \bar{w}_0)$ angehören. Damit ist (39) und zugleich $v_0=\bar{w}_0$ bewiesen.

Die folgenden Betrachtungen werden unter der Voraussetzung (12) ausgeführt. Sie hat $\gamma_1\left(\frac{t}{4}\right)=0$, also $\gamma_0(t)\neq 0$ zur Folge. Wegen $d=-4t$ ist $\varepsilon(*)\left(\frac{-t}{*}\right)$ ein eigentlicher Charakter mod $4t$. Die Voraussetzungen des Haupttheorems von Shimura [8, S. 458] sind daher für c_0 erfüllt, so daß

$$F_t(s)=\sum_{n=1}^{\infty} A_t(n)\,e^{2\pi i n z},$$

definiert durch

$$\sum_{n=1}^{\infty} A_t(n)\, n^{-s} = \sum_{m=1}^{\infty} \varepsilon(m) \left(\frac{-t}{m}\right) m^{k-2-s} \sum_{n=1}^{\infty} \gamma_0(t\, n^2)\, n^{-s}$$

$$= \prod_{p>2} \left(1 - \left(\frac{-t}{m}\right) p^{k-2-s}\right)^{-1} D_t(s) = \gamma_0(t)\, D(s),$$

eine Spitzenform in $(\Gamma_0(2^a), 2k-2)$ mit geeignetem $a \in \mathbb{N}$ darstellt. $D(s)$ erweist sich damit als eine ganze Funktion, woraus erhellt, daß $D(s)$ das Eulerprodukt einer Eigenfunktion $f \in (\Gamma, 2k-2)_0$ ist.

Die Herleitung von (9) aus (10) unter Benutzung von

$$\Psi(s, \zeta) = \pi^{-\frac{s}{2}} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \Psi(1-s, \zeta)$$

darf dem Leser überlassen bleiben.

Literatur

1. Andrianov, A.N.: Eulerprodukte, die den Siegelschen Modulformen zweiten Grades entsprechen (russisch). Uspehi Matematičeskih Nauk **XXIX**, 43–110 (1974)
2. Kurokawa, N.: Examples of eigenvalues of Hecke operators on Siegel cusp forms of degree two. Inv. Math. **49**, 149–165 (1978)
3. Hecke, E.: Über die Bestimmung Dirichletscher Reihen durch ihre Funktionalgleichung. Math. Ann. **112**, 664–699 (1936)
4. Hecke, E.: Über Modulfunktionen und die Dirichletschen Reihen mit Eulerscher Produktentwicklung II. Math. Ann. **114**, 316–351 (1937)
5. Maaß, H.: Lineare Relationen für die Fourierkoeffizienten einiger Modulformen zweiten Grades. Math. Ann. **232**, 163–175 (1978)
6. Maaß, H.: Über eine Spezialschar von Modulformen zweiten Grades. Inv. Math. **52**, 95–104 (1979)
7. Maaß, H.: Über eine Spezialschar von Modulformen zweiten Grades (II). Inv. Math. 53, 249–253 (1979)
8. Shimura, G.: On modular forms of half integral weight. Ann. Math. **97**, 440–481 (1973)

Eingegangen am 18. Juni 1979

Inventiones math. 53, 267–280 (1979)

Inventiones mathematicae

Modular Descent and the Saito-Kurokawa Conjecture*

Anatolii N. Andrianov

Mathematical Institute, Fontanka 27, 191011 Leningrad, USSR

Introduction

Let $\mathfrak{M}_k^2$ be the space of all holomorphic Siegel modular forms of integral weight $k>0$ with respect to the Siegel modular group of degree two $\Gamma_2 = Sp_2(\mathbb{Z})$. Denote by M_k the subspace of all cusp forms

$$F(Z) = \sum_{N\in\mathfrak{N}} f(N) \exp(2\pi i \operatorname{Tr}(NZ)) \tag{1}$$

from $\mathfrak{M}_k^2$, where N runs through the set

$$\mathfrak{N} = \left\{ \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}; \ a, b, c \in \mathbb{Z},\ a>0,\ b^2-4ac<0 \right\},$$

$$Z \in H_2 = \{Z = X+iY \in M_2(\mathbb{C});\ {}^tZ = Z,\ Y>0\},$$

and Tr is the trace, which for every $\begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} \in \mathfrak{N}$ satisfy the following Maaß condition [1]:

$$f\begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} = \sum_{\substack{g \mid a, b, c \\ g>0}} g^{k-1} f\begin{pmatrix} 1 & \frac{b}{2g} \\ \frac{b}{2g} & \frac{ac}{g^2} \end{pmatrix}. \tag{2}$$

It was recently proved by Maaß [2] that

$$\dim M_k = \left[\frac{k-4}{6}\right] \quad \text{if } k \equiv 0 \pmod 2,\ k \geqq 4. \tag{3}$$

The first result of our paper is the following

---

* Dedicated to Hans Maaß

0020-9910/79/0053/0267/$02.80

Theorem 1. *The Maaß space M_k is invariant with respect to all Hecke operators which act on the space $\mathfrak{M}_k^2$.*

It follows from the theorem and the general theory of Hecke operators (see, for example, [3], § 1.3) that there exists a basis $F_1, F_2, \ldots, F_h$ of M_k such that each F_i is an eigenfunction of all Hecke operators. Let F be one of these eigenfunctions. According to [3], § 1.3 one can associate with F the zeta-function $Z_F(s)$ of F which is an Euler product of the form

$$Z_F(s) = \prod_{p \text{ primes}} Q_{p,F}(p^{-s})^{-1}, \tag{4}$$

where $\mathrm{Re}(s)$ is large enough and where for each prime p the function $Q_{p,F}(t)$ is a polynomial in t of degree four whose coefficients are expressed in terms of eigenvalues of Hecke operators $T(p)$ and $T(p^2)$ associated with the eigenfunction F.

Our main result is the following

Theorem 2. *Let $F \in M_k$, $F \not\equiv 0$, be an eigenfunction of all Hecke operators and $Z_F(s)$ the corresponding zeta-function* (4). *Then for each prime number p the polynomial $Q_{p,F}(t)$ has a decomposition of the form*

$$Q_{p,F}(t) = (1 - p^{k-2}t)(1 - p^{k-1}t)(1 - \omega(p)t + p^{2k-3}t^2). \tag{5}$$

If we define $\omega(n)$ for $n = 1, 2, \ldots$ by

$$\sum_{n=1}^{\infty} \omega(n) n^{-s} = \prod_{p \text{ primes}} (1 - \omega(p) p^{-s} + p^{2k-3-2s})^{-1},$$

the series

$$\varphi(z) = \sum_{n=1}^{\infty} \omega(n) \exp(2\pi i n z) \qquad (z = x + iy,\ y > 0)$$

is a cusp form of weight $2k-2$ with respect to the ordinary modular group $\Gamma_1 = SL_2(\mathbb{Z})$.

A similar result with some restrictions was originally obtained by Maaß [4].

It follows from the definition that the modular form $\varphi(z)$ which was obtained in the Theorem 2 by a "descent" from the Siegel eigenfunction F of all Hecke operators is again an eigenfunction of all Hecke operators on the space $\mathfrak{M}_{2k-2}^1$ of all modular forms of weight $2k-2$ with respect to Γ_1. The result (3) of Maaß shows that if $k \equiv 0 \pmod 2$ and $k \geqq 4$ the space M_k has the same dimension as the space of all cusp forms from $\mathfrak{M}_{2k-2}^1$. Therefore if one could prove, for example, that any non zero modular form from M_k which is an eigenfunction of all Hecke operators is uniquely determined up to a constant factor by the corresponding eigenvalues (the analogous statement is true for the space $\mathfrak{M}_{2k-2}^1$; it is perhaps true for the space $\mathfrak{M}_k^2$ or at least for the space M_k but at the moment this is not known) it would follow from Theorem 2 that each $\varphi \in \mathfrak{M}_{2k-2}^1$ which is a cusp form and is an eigenfunction of all Hecke operators can be obtained by the descent or, in other words, that each such φ can be "lifted" to an eigenfunction

$F \in M_k$. The last statement (without an intrinsic description of the space M_k) was communicated to the author by Hiroshi Saito in the beginning of 1977 as a conjecture. Later it was independently communicated by N. Kurokawa and finally it was published by him in more detail in [5].

This work has arisen during the author's stay at Heidelberg University in the spring and summer of 1979 as a result of many discussions of the topic with Professor Hans Maaß, the result of attempts to answer his questions and to continue his work [4], where the first and very considerable progress in the direction of the Saito-Kurokawa conjecture was obtained. The author would like to express to Professor Maaß his deep gratitude.

§1. Invariance of Maaß Spaces under Hecke Operators

Here we shall prove Theorem 1.

Lemma 1. *Let $f\colon \mathfrak{N} \to \mathbb{C}$ be a function on $\mathfrak{N}$ which satisfies*

$$f({}^tUNU) = f(N) \quad \text{for all} \quad N \in \mathfrak{N} \quad \text{and} \quad U \in SL_2(\mathbb{Z}) \tag{6}$$

then f satisfies (2) *if and only if the following two conditions are fulfilled:*

1. $f\begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}$ *depends only on the greatest common divisor $t=(a,b,c)$ of a, b, c and the discriminant b^2-4ac. In other words there is a function $f_0(t,D)$ where $t=1,2,\dots$ and $D<0$, $D \equiv 0, 1 \pmod 4$* such that

$$f\begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} = f_0(t,D) \quad \text{if} \quad (a,b,c)=t \quad \text{and} \quad b^2-4ac=Dt^2. \tag{7}$$

2. *The function $f_0(t,D)$ in* (7) *satisfies*

$$f_0(t,D) = \sum_{g \mid t,\, g>0} g^{k-1} f_0\left(1, \frac{t^2 D}{g^2}\right) \tag{8}$$

for every admissible t and D.

Proof. Let f satisfy (2) and (6). It follows from (2) that it is enough to prove that

$$f\begin{pmatrix} 1 & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} = f\begin{pmatrix} 1 & \frac{b'}{2} \\ \frac{b'}{2} & c' \end{pmatrix}$$

if $b^2-4c = b'^2-4c'$. The last condition implies that there is an integer l such that $\frac{b}{2}+l=\frac{b'}{2}$. By (6) and $b^2-4c=b'^2-4c'$ we get

$$f\begin{pmatrix}1 & \frac{b}{2}\\ \frac{b}{2} & c\end{pmatrix}=f\left(\begin{pmatrix}1 & 0\\ l & 1\end{pmatrix}\begin{pmatrix}1 & \frac{b}{2}\\ \frac{b}{2} & c\end{pmatrix}\begin{pmatrix}1 & l\\ 0 & 1\end{pmatrix}\right)=f\begin{pmatrix}1 & \frac{b'}{2}\\ \frac{b'}{2} & c'\end{pmatrix}.$$

All other statements are obvious. The lemma is proved.

Remark. The coefficients $f(N)$ of the series (1) for $F\in\mathfrak{M}_k^2$ satisfy

$$f({}^tUNU)=(\det U)^k f(N) \qquad (N\in\mathfrak{N},\ U\in GL_2(\mathbb{Z})).$$

Therefore it follows from (7) that

$$M_k=\{0\} \qquad k\equiv 1 \pmod 2). \tag{9}$$

Lemma 2. *A function* $f(t, D)$ $(t=1, 2, \dots, D<0,\ D\equiv 0, 1 \pmod 4))$ *satisfies* (8) *if and only if for any prime number* q, $d=0, 1, 2, \dots, t=1, 2, \dots, (t, q)=1$ *and* D *it satisfies*

$$f(q^d t, D)=\sum_{h=0}^{d} q^{h(k-1)} f(t, q^{2(d-h)} D). \tag{10}$$

The proof is obvious.

It is clear that if q, t, D are fixed the set of conditions (10) for $d=0, 1, \dots$ is equivalent to the following identity between formal power series

$$\sum_{d=0}^{\infty} f(q^d t, D)\, z^d$$
$$=(1-q^{k-1}z)^{-1}\sum_{d=0}^{\infty} f(t, q^{2d} D)\, z^d=(1-q^{k-1}z)^{-1}[f]_q(t, D, z). \tag{11}$$

Proof of Theorem 1. The ring L of all Hecke operators on $\mathfrak{M}_k^2$ is generated by its p-components L_p for all prime numbers p and each operator from L_p for a prime p is a polynomial in two operators $T(p)$ and $T(p^2)$ with constant coefficients (see [3], §1.3). Therefore it is sufficient to prove that for each prime p the space M_k is invariant with respect to the operators $T(p)$ and $T(p^2)$. It will be more convenient for us to consider $T'(p)=T(p)^2-T(p^2)$ in place of $T(p^2)$.

Let F be a function from M_k with Fourier expansion (1). Denote by $g(N)$ and $g'(N)(N\in\mathfrak{N})$ the Fourier coefficients of the modular forms $T(p)F$ and $T'(p)F$ respectively, where p is a fixed prime number. It follows from the proposition [3] 2.1.2 and the formula [6] (3.15) that the functions $g(N)$ and $g'(N)$ can be written in the form

$$g(N)=(\Delta^+(p)f)(N)+p^{k-2}(\Pi(p)f)(N)+p^{2k-3}(\Delta^-(p)f)(N),$$
$$g'(N)=p^{2k-4}(v_p f)(N)+p^{2k-3}(\Delta^+(p)\,\Delta^-(p)\,f)(N)$$
$$+p^{3k-5}(\Pi(p)\Delta^-(p)f)(N)+p^{k-2}(\Delta^+(p)\Pi(p)f)(N), \tag{12}$$

where $\Delta^+(p)$, $\Delta^-(p)$, $\Pi(p)$ are the operators which were defined in [3] §2.1 and v_p

is defined by $(v_p f)(N)=v_p(N)\, f(N)$, where $v_p(N)$ for $N=\begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}$ is the number of solutions of the congruence $ax^2+bxy+cy^2\equiv 0 \pmod p$ on the projective line mod p.

Let $f_0(t, D)$ be the function associated with f by (7) and let $N\in\mathfrak{N}$. We can write $N=p^d t \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}$, where $d\geqq 0$, $(t, p)=1$, $(a, b, c)=1$. We shall also write $D=b^2-4ac=D_0 l^2$, where D_0 is the discriminant of the imaginary quadratic field $\mathbb{Q}(\sqrt{D})$. It follows from (12), the definitions of the operators $\Delta^+(p)$, $\Delta^-(p)$, $\Pi(p)$, v_p and the formulae for the action of $\Pi(p)$ which were established in the theorems [3] 2.3.1 and 2.3.2 that the functions $g(N)$ and $g'(N)$ can be expressed in the form

$$\begin{aligned} g(N)=&f_0(p^{d+1}t, D)+p^{2k-3}f_0(p^{d-1}t, D)\\ &+p^{k-2}\, v f_0(p^d t, D)+p^{k-2}(p+1-v)\, f_0(p^{d-1}t, p^2 D) \end{aligned} \tag{13}$$

if $l\not\equiv 0 \pmod p$, where $v=v_p(p^{-d}t^{-1}N)$ and we suppose here and in the following that $f(t', D')=0$ if t' is not an integral number;

$$\begin{aligned} g(N)=&f_0(p^{d+1}t, D)+p^{2k-3}f_0(p^{d-1}t, D)\\ &+p^{k-2}f_0(p^{d+1}t, p^{-2}D)+p^{k-1}f_0(p^{d-1}t, p^2 D) \end{aligned} \tag{14}$$

if $l\equiv 0 \pmod p$; further

$$\begin{aligned} g'(N)=&p^{2k-4}\, v_p(N)\, f_0(p^d t, D)+p^{2k-3}f_0(p^d t, D)\\ &+p^{3k-5}\, v f_0(p^{d-1}t, D)+p^{3k-5}(p+1-v)\, f_0(p^{d-2}t, p^2 D)\\ &+p^{k-2}\, v f_0(p^{d+1}t, D)+p^{k-2}(p+1-v)\, f_0(p^d t, p^2 D) \end{aligned} \tag{15}$$

if $l\not\equiv 0 \pmod p$; and

$$\begin{aligned} g'(N)=&p^{2k-4}\, v_p(N)\, f_0(p^d t, D)+p^{2k-3}f_0(p^d t, D)\\ &+p^{3k-5}f_0(p^d t, p^{-2}D)+p^{3k-4}f_0(p^{d-2}t, p^2 D)\\ &+p^{k-2}f_0(p^{d+2}t, p^{-2}D)+p^{k-1}f_0(p^d t, p^2 D) \end{aligned} \tag{16}$$

if $l\equiv 0 \pmod p$. It is easy to see that $v_p(N)=p+1$ if $d\geqq 1$ and $v_p(N)=v$ if $d=0$. Besides, v obviously depends only on D. Therefore the formulae (13)–(16) show that the functions $g(N)$, $g'(N)$ satisfy the condition 1 of Lemma 1. Then Lemma 1 shows that it is sufficient to prove that the functions g_0 and g'_0 associated with g and g' by (7) satisfy (8), or, by Lemma 2, (10) for any prime number q, $d\geqq 0$, $(t, q)=1$, and D. Since f_0 satisfies (10) and the relations (10) are linear with respect to f the formulae (13)–(14) for $g_0(p^d t, D)$ and (15)–(16) for $g'_0(p^d t, D)$ show that g_0 and g'_0 satisfy (10) for all prime numbers $q\neq p$. With the formulae (13)–(16) to check whether g_0 and g'_0 satisfy (10) or (11) for $q=p$ is nothing but routine. For example let us prove the relation (11) for g_0' in the case when $l \not\equiv 0 \pmod p$. From (15) we obtain

$$\begin{aligned}
&\sum_{d=0}^{\infty} g_0'(p^d t, D) z^d \\
&\quad = p^{2k-4}(v-p-1) f_0(t,D) + p^{2k-4}(p+1) \sum_{d=0}^{\infty} f_0(p^d t, D) z^d \\
&\qquad + p^{2k-3} \sum_{d=0}^{\infty} f_0(p^d t, D) z^d + p^{3k-5} vz \sum_{d=0}^{\infty} f_0(p^d t, D) z^d \\
&\qquad + p^{3k-5}(p+1-v) z^2 \sum_{d=0}^{\infty} f_0(p^d t, p^2 D) z^d \\
&\qquad + p^{k-2} vz^{-1} \sum_{d=1}^{\infty} f_0(p^d t, D) z^d \\
&\qquad + p^{k-2}(p+1-v) \sum_{d=0}^{\infty} f_0(p^d t, p^2 D) z^d.
\end{aligned}$$

Since f_0 satisfies (11) the last expression is equal to

$$\begin{aligned}
&p^{2k-4}(v-p-1) f_0(t,D) + p^{2k-4}(p+1) H^{-1}[f_0] \\
&\qquad + p^{2k-3} H^{-1}[f_0] + p^{3k-5} vz H^{-1}[f_0] \\
&\qquad + p^{3k-5}(p+1-v) z^2 H^{-1}([f_0] - f_0(t,D)) z^{-1} \\
&\qquad + p^{k-2} vz^{-1}(H^{-1}[f_0] - f_0(t,D)) \\
&\qquad + p^{k-2}(p+1-v) H^{-1}([f_0] - f_0(t,D)) z^{-1},
\end{aligned}$$

where $H = (1 - p^{k-1} z)$, $[f_0] = [f_0]_p(t, D, z)$ (see (11)). On the other hand using (15) for $d=0$ and (16) for $d \geqq 1$ we obtain

$$\begin{aligned}
&H^{-1} \sum_{d=0}^{\infty} g_0'(t, p^{2d} D) z^d \\
&\quad = H^{-1} \Big\{ p^{2k-4} v f_0(t,D) + p^{2k-3} f_0(t,D) \\
&\qquad + p^{k-2} v f_0(pt, D) + p^{k-2}(p+1-v) f_0(t, p^2 D) \\
&\qquad + p^{2k-4} \sum_{d=1}^{\infty} f_0(t, p^{2d} D) z^d + p^{2k-3} \sum_{d=1}^{\infty} f_0(t, p^{2d} D) z^d \\
&\qquad + p^{3k-5} z \sum_{d=0}^{\infty} f_0(t, p^{2d} D) z^d + p^{k-2} z \sum_{d=0}^{\infty} f_0(tp^2, p^{2d} D) z^d \\
&\qquad + p^{k-1} z^{-1} \sum_{d=2}^{\infty} f_0(t, p^{2d} D) z^d \Big\}.
\end{aligned}$$

It follows from (10) that

$$\begin{aligned}
&f_0(pt, p^{2d} D) = f_0(t, p^{2(d+1)} D) + p^{k-1} f_0(t, p^{2d} D), \\
&f_0(p^2 t, p^{2d} D) = f_0(t, p^{2(d+2)} D) + p^{k-1} f_0(t, p^{2(d+1)} D) + p^{2k-2} f(t, p^{2d} D).
\end{aligned}$$

Therefore the last expression can be written in the form

$$\begin{aligned}
H^{-1}\{&p^{2k-4}\, v f_0(t,D)+p^{2k-3} f_0(t,D)\\
&+p^{k-2}\, v(f_0(t,p^2D)+p^{k-1} f_0(t,D))+p^{k-2}(p+1-v)\, f_0(t,p^2D)\\
&+p^{2k-4}([f_0]-f_0(t,D))+p^{2k-3}([f_0]-f_0(t,D))\\
&+p^{3k-5}\, z[f_0]+p^{k-2}\, z^{-1}([f_0]-f_0(t,D)-f_0(t,p^2D)\, z)\\
&+p^{2k-3}([f_0]-f_0(t,D))+p^{3k-4}\, z[f_0]\\
&+p^{k-1}\, z^{-1}([f_0]-f_0(t,D)-f_0(t,p^2D)\, z)\}.
\end{aligned}$$

It is easy to see that the expressions we have obtained are equal. The other cases are similar. Theorem 1 is proved.

§ 2. Modular Descent

Proof of Theorem 2. It follows from (9) that we can suppose that k is even. Let us consider the function

$$\begin{aligned}
\Psi(s)&=(2\pi)^{-s}\,\Gamma(s)\,\zeta(s-k+2)^{-1}\,\zeta(s-k+1)^{-1}\,Z_F(s)\\
&=(2\pi)^{-s}\,\Gamma(s)\prod_{p\ \text{primes}}(1-p^{k-2-s})(1-p^{k-1-s})\,Q_{p,F}(p^{-s})^{-1},
\end{aligned}\tag{17}$$

where $\Gamma(s)$ is the gamma-function and $\zeta(s)$ is the Riemann zeta-function. If $\mathrm{Re}(s)$ is large enough the infinite product in (17) converges absolutely and uniformily and therefore it defines in some half-plane a holomorphic function of s (see [3], § 1.3).

Lemma 3. *The function $\Psi(s)$ can be meromorphically continued to the entire s-plane and satisfies the functional equation*

$$\Psi(2k-2-s)=-\Psi(s).$$

Proof. Let us write $\Psi(s)=\Psi_1(s)\,\eta(s)^{-1}$, where $\Psi_1(s)=(2\pi)^{-2s}\,\Gamma(s)\,\Gamma(s-k+2)\,Z_F(s)$ and $\eta(s)=(2\pi)^{-s}\,\Gamma(s-k+2)\,\zeta(s-k+1)\,\zeta(s-k+2)$. According to the theorem [3] 3.1.1 the function $\Psi_1(s)$ can be meromorphically continued to the entire s-plane and satisfies the functional equation $\Psi_1(2k-2-s)=(-1)^k\,\Psi_1(s)=\Psi_1(s)$. It follows from the well-known properties of the Riemann zeta-function that the same statement is true for $\eta(s)$ with the functional equation $\eta(2k-2-s)=-\eta(s)$. The lemma is proved.

Now we start to prove that $\Psi(s)$ is in fact an entire function. Let us consider the function

$$\Theta_1(z,u)=\sum_{\begin{pmatrix}1 & \frac{b}{2}\\ \frac{b}{2} & c\end{pmatrix}\in\mathfrak{R}} f\begin{pmatrix}1 & \frac{b}{2}\\ \frac{b}{2} & c\end{pmatrix}\exp(2\pi i(zc+ub)),\tag{18}$$

where $z,u\in\mathbb{C}$, $\mathrm{Im}(z)>0$, and $f(N)$ are the Fourier coefficients of F (see (1)). It was proved by Eichler [7] that

$$\Theta_1(z,u)=c_0(z)\,\vartheta_0(z,u)+c_1(z)\,\vartheta_1(z,u), \tag{19}$$

where the functions c_0, c_1 are the components of a holomorphic vector modular form $c=\begin{pmatrix}c_0\\ c_1\end{pmatrix}$ of weight $k-\frac{1}{2}$ with respect to the group $\Gamma_1=SL_2(\mathbb{Z})$ (i.e. $c(z)$ is holomorphic on the upper halfplane including the point $i\infty$ and for each matrix $M=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\in\Gamma_1$ it satisfies the functional equation

$$c((\alpha z+\beta)(\gamma z+\delta)^{-1})=(\gamma z+\delta)^{k-\frac{1}{2}}\,\varepsilon(M)\,c(z), \tag{20}$$

where $\varepsilon(M)$ is an unitary matrix of order two, and for $h=0,1$

$$\vartheta_h(z,u)=\sum_{n=-\infty}^{+\infty}\exp\left(2\pi i\left(z\left(n+\frac{h}{2}\right)^2+2\left(n+\frac{h}{2}\right)u\right)\right)$$

are the theta series.

For $z\in\mathbb{C}$, we define $z^{\frac{1}{2}}$ so that $-\frac{\pi}{2}<\arg z^{\frac{1}{2}}\leqq\frac{\pi}{2}$ and we put $z^{\frac{1}{2}m}=(z^{\frac{1}{2}})^m$ for every $m\in\mathbb{Z}$. It follows from the well-known transformation formulae of the theta series that the vector series $\vartheta(z)=\begin{pmatrix}\vartheta_0(z,0)\\ \vartheta_1(z,0)\end{pmatrix}$ satisfies

$$\vartheta((\alpha z+\beta)(\gamma z+\delta)^{-1})=(\gamma z+\delta)^{\frac{1}{2}}\,\eta(M)\,\vartheta(z), \tag{21}$$

for every $M=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\in\Gamma_1$, where $\eta(M)$ is a unitary matrix of order two, which for the generators of Γ_1 is given by

$$\eta\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix},$$

$$\eta\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}=\exp\left(-\frac{\pi i}{4}\right)\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}.$$

It was shown by Eichler [7] (see also [1], §2) that

$$\varepsilon(M)=\overline{\eta(M)}\qquad (M\in\Gamma_1). \tag{22}$$

It follows from (18) and (19) that

$$c_0(z)=\sum_{c=1}^{\infty}f\begin{pmatrix}1 & 0\\ 0 & c\end{pmatrix}\exp(2\pi icz),$$

$$c_1(z)=\sum_{c=1}^{\infty}f\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & c\end{pmatrix}\exp(2\pi i(c-\tfrac{1}{4})z). \tag{23}$$

Lemma 4. *Under the conditions of the Theorem 2 there exists an integer* $c\geqq 1$ *such that* $f\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & c\end{pmatrix}\neq 0$ *and the number* $d=4c-1$ *is not divisible by the square* l^2 *of an integer* $l>1$, *so that the discriminant of the quadratic field* $\mathbb{Q}(\sqrt{-d})$ *coincides with* $-d$.

Proof. It was proved by Maaß [1] that a modular form $F \in M_k$ is uniquely determined by the series $\Theta_1(z,u)$. On the other hand it follows from (20) that $c_0(z)$ is uniquely determined by $c_1(z)$. Therefore $c_1(z) \neq 0$. Let us consider the smallest c such that $f\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & c \end{pmatrix} \neq 0$. Suppose that $4c-1=dl^2$ where $-d$ is the discriminant of the field $\mathbb{Q}(\sqrt{1-4c})$ and $l>1$. We shall obtain a contradiction using the following lemma which will be proved below.

Lemma 5. *If $F \in M_k$ is an eigenfunction of all Hecke operators and $-d \equiv 1 \pmod 4$ is the discriminant of an imaginary quadratic field K, then the following identity is valid*

$$\zeta_K(s-k+2)\,\zeta(s-k+1)\sum_{b=1}^{\infty} b^{-s} f\begin{pmatrix} 1 & \frac{b}{2} \\ \frac{b}{2} & \frac{b^2(d+1)}{4} \end{pmatrix}$$
$$= f\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{(d+1)}{4} \end{pmatrix} Z_F(s), \tag{24}$$

where $\mathrm{Re}(s)$ *is large enough, $f(N)$ are the Fourier coefficients of F, ζ_K is the Dedekind zeta-function of the field K, ζ is the Riemann zeta-function and Z_F is the zeta-function* (4).

Let us finish the proof of Lemma 4. Since $l>1$, we have

$$\frac{d+1}{4} < \frac{(dl^2+1)}{4}.$$

Therefore $f\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{(d+1)}{4} \end{pmatrix} = 0$. Then the identity (24) shows that

$$f\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & c \end{pmatrix} = f\left(\begin{pmatrix} 1 & 0 \\ \frac{(l-1)}{2} & 1 \end{pmatrix}\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & c \end{pmatrix}\begin{pmatrix} 1 & \frac{(l-1)}{2} \\ 0 & 1 \end{pmatrix}\right)$$
$$= f\begin{pmatrix} 1 & \frac{l}{2} \\ \frac{l}{2} & \frac{l^2(d+1)}{4} \end{pmatrix} = 0.$$

The contradiction proves Lemma 4, if Lemma 5 is proved.

Proof of Lemma 5. From the identity of Theorem [3] 2.4.1 with $\chi \equiv 1$ and the property (7) of the function f we get the identity

$$\zeta_K(s-k+2)\sum_{m=1}^{\infty} f(mN_0)\,m^{-s} = f(N_0)\,Z_F(s),$$

where $N_0=\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{(d+1)}{4}\end{pmatrix}$. It follows from (2) that

$$\sum_{m=1}^{\infty} f(mN_0)\,m^{-s}$$

$$=\sum_{m=1}^{\infty}\left(\sum_{m=ab} a^{k-1}(ab)^{-s} f\begin{pmatrix}1 & \frac{b}{2}\\ \frac{b}{2} & \frac{b^2(d+1)}{4}\end{pmatrix}\right)$$

$$=\zeta(s-k+1)\sum_{b=1}^{\infty} b^{-s} f\begin{pmatrix}1 & \frac{b}{2}\\ \frac{b}{2} & \frac{b^2(d+1)}{4}\end{pmatrix}.$$

The lemma follows from the identities.

Lemma 6. *Let* $F\in M_k$, $d>1$, $d\equiv 3 \pmod 4$, *then the following identity is valid*

$$\sum_{b=1}^{\infty} b^{-s} f\begin{pmatrix}1 & \frac{b}{2}\\ \frac{b}{2} & \frac{b^2(d+1)}{4}\end{pmatrix}$$

$$=(\pi d)^{\frac{1}{2}s}\left\{2\Gamma\left(\frac{s}{2}\right)\right\}^{-1}\int_0^{\infty}\left(\int_0^1 (c_0(z)\overline{\vartheta_0(dz)}+c_1(z)\overline{\vartheta_1(dz)})\,y^{\frac{1}{2}s-1}\,dx\right)dy, \tag{25}$$

where $\operatorname{Re}(s)$ *is large enough,* $z=x+iy$, c_0, c_1 *are determined by* (18)–(19), $\vartheta_0(z)=\vartheta_0(z,0)$, $\vartheta_1(z)=\vartheta_1(z,0)$, *and the bar means complex conjugation.*

Proof. It immediately follows from (23) and the definition of the theta series ϑ_0, ϑ_1 that

$$\int_0^{\infty}\left(\int_0^1 (c_0(z)\overline{\vartheta_0(dz)}+c_1(z)\overline{\vartheta_1(dz)})\,y^{\frac{1}{2}s-1}\,dx\right)dy$$

$$=2\Gamma\left(\frac{s}{2}\right)(\pi d)^{-\frac{1}{2}s}\left(\sum_{n=1}^{\infty}(2n)^{-s} f\begin{pmatrix}1 & 0\\ 0 & dn^2\end{pmatrix}\right.$$

$$\left.+\sum_{n=0}^{\infty}(2n+1)^{-s} f\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{(d(2n+1)^2+1)}{4}\end{pmatrix}\right).$$

Since

$$f\begin{pmatrix}1 & 0\\ 0 & dn^2\end{pmatrix}=f\left(\begin{pmatrix}1 & 0\\ n & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & dn^2\end{pmatrix}\begin{pmatrix}1 & n\\ 0 & 1\end{pmatrix}\right)$$
$$=f\begin{pmatrix}1 & \frac{2n}{2}\\ \frac{2n}{2} & \frac{(2n)^2(d+1)}{4}\end{pmatrix}$$

and similarly

$$f\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{(d(2n+1)^2+1)}{4}\end{pmatrix}=f\begin{pmatrix}1 & \frac{(2n+1)}{2}\\ \frac{(2n+1)}{2} & \frac{(2n+1)^2(d+1)}{4}\end{pmatrix}.$$

The lemma is proved.

To transform the integral from Lemma 6 we shall need the following lemma which may be interesting in itself.

Lemma 7. *Let* $d>1$, $d\equiv 3 \pmod 4$,

$$M=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\in\Gamma_1,\qquad \gamma\equiv 0 \pmod d,\qquad \tilde{M}=\begin{pmatrix}\alpha & d\beta\\ \frac{\gamma}{d} & \delta\end{pmatrix}$$

then

$$\eta(\tilde{M})=\chi_{-d}(\delta)\,\overline{\eta(M)}, \tag{26}$$

where $\eta(M)$ $(M\in\Gamma_1)$ *are the multiplier matrices of* (21) *and* χ_{-d} *is a Dirichlet character* mod d *which can be defined by*

$$\chi_{-d}(\delta)=(\operatorname{sign}\delta)\left(\frac{-d}{|\delta|}\right), \tag{27}$$

where $(-)$ *is the generalised Legendre symbol.*

Proof. Let us consider the theta series

$$\Theta_H(z)=\sum_{m,n=-\infty}^{\infty}\exp(2\pi i z H(m,n))$$

of the quadratic form $H(x,y)=x^2+xy+\frac{1+d}{4}y^2$. The level of H is d and its discriminant is $-d$. Therefore, as is well known, $\Theta_H(z)$ satisfies

$$\Theta_H((\alpha z+\beta)(\gamma z+\delta)^{-1})=\chi_{-d}(\delta)(\gamma z+\delta)\,\Theta_H(z).$$

On the other hand, one can easily check that

$$\Theta_H(z)=\vartheta_0(z)\,\vartheta_0(dz)+\vartheta_1(z)\,\vartheta_1(dz)={}^t\vartheta(z)\,\vartheta(dz),$$

where $\vartheta(z)=\begin{pmatrix}\vartheta_0(z,0)\\ \vartheta_1(z,0)\end{pmatrix}$. Therefore it follows from (21) that

$$\begin{aligned}&\Theta_H((\alpha z+\beta)(\gamma z+\delta)^{-1})\\ &\quad={}^t\vartheta((\alpha z+\beta)(\gamma z+\delta)^{-1})\,\vartheta\left((\alpha dz+d\beta)\left(\left(\frac{\gamma}{d}\right)dz+\delta\right)^{-1}\right)\\ &\quad=(\gamma z+\delta)\,{}^t\vartheta(z)\,{}^t\eta(M)\,\eta(\tilde{M})\,\vartheta(dz).\end{aligned}$$

It is easy to see that an identity ${}^t\vartheta(z)\,A\,\vartheta(dz)=0$ with a constant matrix A and $d>1$ is possible only if $A=0$. Therefore from the relations we have obtained we get that ${}^t\eta(M)\,\eta(\tilde{M})=\chi_{-d}(\delta)\,E_2$. The lemma is proved.

Let us come back to the proof of Theorem 2. Let c satisfy the conditions of Lemma 4 and $d=4c-1$. Then by (17) and (24) we get

$$a\Psi(s)=(2\pi)^{-s}\,\Gamma(s)\,\zeta_K(s-k+2)\,\zeta(s-k+2)^{-1}\sum_{b=1}^{\infty}b^{-s}f\begin{pmatrix}1 & b/2\\ b/2 & b^2c\end{pmatrix},$$

where $a=f\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & c\end{pmatrix}\neq 0$ and ζ_K is the Dedekind zeta-function of the field $K=\mathbb{Q}(\sqrt{-d})$. It is well known that $\zeta_K(s)=\zeta(s)\,L(s,\chi_{-d})$, where $L(s,\chi_{-d})$ is the Dirichlet L-series with the character (27). Using (25) and the formula

$$\sqrt{\pi}\,\Gamma(s)=2^{s-1}\,\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right)$$

we obtain

$$a\,\Psi(s)=2^{-2}\,\pi^{-\frac{1}{2}(s+1)}\,d^{\frac{1}{2}s}\,\Gamma\left(\frac{s+1}{2}\right)L(s-k+2,\chi_{-d})\int_S G(z)\,y^{\frac{1}{2}s+1}\,dz, \tag{28}$$

where $G(z)={}^tc(z)\,\overline{\vartheta(dz)}$, $\quad dz=y^{-2}\,dx\,dy$ and

$$S=\{x+iy;\ 0\leqq x\leqq 1,\ y>0\}.$$

Using the standard "Rankin trick" we can write

$$S=\bigcup_{M\in\Gamma_0\backslash\Gamma_0(d)}M\cdot(D_0(d)),$$

where

$$\Gamma_0(d)=\left\{\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\in\Gamma_1,\ \gamma\equiv 0\ (\mathrm{mod}\ d)\right\},\qquad \Gamma_0=\left\{\begin{pmatrix}\pm 1 & t\\ 0 & \pm 1\end{pmatrix}\in\Gamma_1\right\}$$

and $D_0(d)$ is a fundamental domain of the group $\Gamma_0(d)$ on the upper half-plane. If

$M=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\in\Gamma_0(d)$ it follows from (21) and (26) that

$$\overline{\vartheta(dM(z))}=\overline{\vartheta\left((\alpha dz+d\beta)\left(\left(\frac{\gamma}{d}\right)dz+\delta\right)^{-1}\right)}$$
$$=\overline{(\gamma z+\delta)^{\frac{1}{2}}}\,\overline{\eta(\tilde{M})}\,\overline{\vartheta(dz)}=\chi_{-d}(\delta)\overline{(\gamma z+\delta)^{\frac{1}{2}}\,\eta(M)}\,\overline{\vartheta(dz)}.$$

It then follows from (20) and (22) that

$$G(M(z))=\chi_{-d}(\delta)(\gamma z+\delta)^{k-1}\,|\gamma z+\delta|\,G(z).$$

Therefore we can write

$$\int_S G(z)\,y^{\frac{1}{2}s+1}\,dz$$
$$=\sum_{M\in\Gamma_0\backslash\Gamma_0(d)}\ \int_{D_0(d)} G(M(z))\,\{y(M(z))\}^{\frac{1}{2}s+1}\,dM(z)$$
$$=\sum_{\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)\in\Gamma_0\backslash\Gamma_0(d)}\ \int_{D_0(d)}\chi_{-d}(\delta)(\gamma z+\delta)^{k-1}\,|\gamma z+\delta|\,G(z)\{y\cdot|\gamma z+\delta|^{-2}\}^{\frac{1}{2}s+1}\,dz$$
$$=\int_{D_0(d)} G(z)\,y^{\frac{1}{2}s+1}\Big\{\sum_{\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)\in\Gamma_0\backslash\Gamma_0(d)}\chi_{-d}(\delta)(\gamma z+\delta)^{k-1}\,|\gamma z+\delta|^{-s-1}\Big\}\,dz.$$

It follows from the known description of $\Gamma_0\backslash\Gamma_0(d)$ (see Lemma [8] 3.2) and (28) that we can finally write

$$\Psi(s)=(8a)^{-1}\,d^{\frac{1}{2}s}\int_{D_0(d)} G(z)\,y^{\frac{1}{2}}\,H_{k-1}\left(\frac{s+1}{2},z,\chi_{-d}\right)dz, \tag{29}$$

where

$$H_t(s,z,\chi)=\pi^{-s}\,\Gamma(s)\,y^s\sum_{m,n}\chi(n)(mdz+n)^t\,|mdz+n|^{-2s}$$

and the summation is taken over all $(m,n)\in\mathbb{Z}^2$ other than $(0,0)$.

Observe that the functions $c_0(z)$ and $c_1(z)$ together with F are cusp forms. Therefore it follows from (29) and Shimura's Lemma [8] 3.3 that the function (29) is holomorphic for all s and bounded in every vertical strip $\sigma_1<\mathrm{Re}(s)<\sigma_2$. From a well-known theorem of Hecke (see, for example, [9]) and Lemma 3 it then follows that

$$\Psi(s)=(2\pi)^{-s}\,\Gamma(s)\sum_{n=1}^{\infty}\omega(n)\,n^{-s},$$

where the function $\varphi(z)=\sum_{n=1}^{\infty}\omega(n)\exp(2\pi i n z)$ is a cusp form of weight $2k-2$ with respect to the group Γ_1. The identity (17) shows that the Dirichlet series

$\sum_{n=1}^{\infty} \omega(n)\, n^{-s}$ has an Euler decomposition. It then follows from another result of Hecke ([10], pp. 687–688) that $\varphi(z)$ is an eigenfunction of all Hecke operators and

$$\sum_{n=1}^{\infty} \omega(n)\, n^{-s} = \prod_{p \text{ primes}} (1-\omega(p)\, p^{-s} + p^{2k-3-2s})^{-1}.$$

Since the Euler decomposition is unique, we get

$$(1-p^{k-2-s})(1-p^{k-1-s})\, Q_{p,F}(p^{-s})^{-1} = (1-\omega(p)\, p^{-s} + p^{2k-3-2s})^{-1}.$$

Theorem 2 is proved.

Since χ_{-d} in (29) is primitive it is perhaps possible to get the functional equation for $\Psi(s)$ directly from (29) and the functional equation for $H_t(s, z, \chi)$.

References

1. Maaß, H.: Über eine Spezialschar von Modulformen zweiten Grades. Inv. Math. **52**, 95–104 (1979)
2. Maaß, H.: Über eine Spezialschar von Modulformen zweiten Grades II. Inv. Math. **53**, 249–253 (1979)
3. Andrianov, A.N.: Euler products associated with Siegel modular forms of degree 2. Uspehi Mat. Nauk **29**, 43–110 (1974)
4. Maaß, H.: Über eine Spezialschar von Modulformen zweiten Grades III. Inv. Math. **53**, 255–265 (1979)
5. Kurokawa, N.: Examples of Eigenvalues of Hecke Operators on Siegel Cusp Forms of Degree Two. Inv. Math. **49**, 149–165 (1978)
6. Andrianov, A.N.: Symmetric squares of zeta functions of Siegel modular forms of degree 2. Trudy Mat. Inst. Steklov **142**, 22–45 (1976)
7. Eichler, M.: Über die Anzahl der linear unabhängigen Siegelschen Modulformen von gegebenem Gewicht. Math. Ann. **213**, 281–291 (1975)
8. Shimura, G.: On modular forms of half integral weight, Ann. of Math. **97**, 440–481 (1973)
9. Ogg, A.: Modular Forms and Dirichlet Series. New York: Benjamin 1969
10. Hecke, E.: Mathematische Werke, Göttingen: Vandenhoeck & Ruprecht 1959, 1970

Received June 21, 1979

Inventiones math. 53, 281–297 (1979)

Inventiones mathematicae

Busemann Functions and Total Curvature

Katsuhiro Shiohama*

Department of Mathematics, University of Tsukuba, 300-31, Ibaraki, Japan

Introduction

Let M^m be a connected, complete and noncompact Riemannian manifold of dimension $m \geqq 2$, which has no boundary. A *ray* is by definition a geodesic parametrized by arc length on $[0, \infty)$ any of whose subarc realizes distance between its end points. It is a nature of noncompactness that through every point p on M^m there passes at least a ray γ: $[0, \infty) \to M$ which is parametrized by arc length. A Busemann function F_γ: $M^m \to R$ with respect to γ is defined as follows: For every point $x \in M^m$ let

$$F_\gamma(x) = \lim_{t \to \infty} [t - d(x, \gamma(t))],$$

where d is by definition the distance function with respect to the Riemannian structure on M^m. The right hand side is bounded from above by $d(p, x)$ and is monotone increasing with t. It converges uniformly on every compact set of M^m as $t \to \infty$. This function was first defined on G-spaces in § 22 of [2] and used to define the parallel axiom on a straight space.

It turns out from Theorem 1.10 in [3] that F_γ is convex if the sectional curvature of M^m is nonnegative everywhere. And the structure of such manifolds has already been investigated in [3, 8, 11] and others. Recently the structure of complete and noncompact manifolds with locally nonconstant convex functions has been investigated in [1, 8, 7] and others. As is seen in Theorem B of [7], a convex function on M^m has such a nice property that if there is a compact level set then so are all other levels. And this fact plays an important role for the proof of results obtained in this article.

Now we are interested in the behavior of Busemann functions on complete noncompact Riemannian manifolds of nonnegative sectional curvature. As is intuitively seen in simple examples of surfaces of revolutions in Euclidean 3-space whose profile curves are paraboloids and two fold hyperboloids, the total curvature will give restrictions to the behavior of Busemann functions on such surfaces. On

* Dedicated to Professor Isamu Mogi on his 60th Birthday

0020-9910/79/0053/0281/$03.40

complete, oriented and noncompact Riemannian 2-manifolds of positive Gaussian curvature, the total curvature $\int_{M^2} G\,d\mu$ is bounded from above by 2π, which is known as a theorem of Cohn-Vossen. Indeed it is stated on p. 80 of [5] that if $G>0$ everywhere on M^2, then the total curvature exists and hence Satz 6 on p. 79 implies that the total curvature of $M^2 \leqq 2\pi\chi(M^2) \leqq 2\pi$. And recently a simple proof for this inequality under weaker assumption $G \geqq 0$ is given in [10].

A continuous function $f\colon M^m \to R$ is said to be *exhaustion* if $f^{-1}((-\infty, a])$ is compact for every $a \in R$. In the examples of surfaces of revolutions, we see that all Busemann functions are exhaustion if the total curvature is roughly speaking "large", and they are non-exhaustion if the total curvature is roughly speaking "small". More generally it will be conjectured that *if there is an exhaustion Busemann function on M^m on nonnegative sectional curvature then so is every Busemann function on it, and similarly, if there is a non-exhaustion Busemann function then so is every Busemann function on it.* This is solved affirmatively in the simplest case where dim $M=2$ (see Corollary to Theorem 4.4). As is summarized at the end of introduction we shall investigate that the behavior of Busemann functions are controled by the total curvature.

As is seen in the examples stated above, metric balls with the same center on M^m will have in common "larger" diameter if a Busemann function is non-exhaustion, and they will have in common "small" diameter if a Busemann function is exhaustion. These phenomena are investigated in §§ 1 and 2. We shall observe in § 1 that the ratio $\delta_p(t)/t$ of the diameter $\delta_p(t)$ of the boundary of metric t-ball around a point $p \in M^2$ has a uniform property. Namely, it turns out from Theorem 1.2 that if limit of $\delta_p(t_j)/t_j$ exists for some point p and for some monotone increasing divergent sequence, then the limit of it is not smaller than the limit sup $\delta_q(t_j)/t_j$ for any $q \in M^2$. And in § 2, we know from Proposition 2.2 and Theorem 2.4 that if a Busemann function $F_\alpha\colon M^m \to R$ is non-exhaustion and if it does not take minimum, then $\delta_p(t)/t > \sqrt{2}$ for all $t>0$, where $p=\alpha(0)$. Moreover if there is a point $x \in M^m$ and if there is a monotone increasing divergent sequence $\{t_j\}$ so that $\delta_x(t_j)/t_j > \sqrt{2}$, then we see with the aid of Proposition 2.1 that $\delta_x(t)/t > \sqrt{2}$ for all $t>0$ and $\delta_y(t)/t \geqq \sqrt{2}$ holds for every point $y \in M^m$. Thus in this case we see that through every point on M^m there passes at least two rays with respect to which Busemann functions are non-exhaustion. Furthermore we see from Theorems 2.4 and 4.2 that $\delta_p(t_j)/t_j > \sqrt{2}$ (or $\leqq \sqrt{2}$ respectively) holds for some monotone increasing divergent sequence and for some point p if and only if every Busemann function on M^2 is non-exhaustion (or exhaustion respectively).

Moreover in the special case where $\dim M=2$ and where there is a non-exhaustion Busemann function which takes minimum, we see in Theorem 3.1 that the total curvature of M^2 is not greater than π. The minimum set of a non-exhaustion Busemann function is a ray if the Gaussian curvature G is nonnegative and not identically zero. And the support of G is contained in a closure compact domain bounded by a geodesic biangle with edge angles $\pi/2$ if and only if the Busemann function with respect to the minimum set takes a minimum. In this case the total curvature of M^2 is exactly π (see Theorem 3.1).

Finally in § 4 we obtain one of our main results which states as follows: Let M^2 be a connected, complete, oriented and noncompact Riemannian 2-manifold without boundary. If the Gaussian curvature G is nonnegative everywhere, then the total curvature o M^2 is not greater than π if and only if no Busemann function is exhaustion, and the total curvature is greater than π if and only if no Busemann function is non-exhaustion (see Theorem 4.4). Indeed we prove in Theorem 4.1 that if there is an exhaustion Busemann function then the total curvature is greater than π. And in Theorem 3.2 we prove that if there is a non-exhaustion Busemann function then the total curvature is not greater than π.

We refer to [4] for basic tools in Riemannian geometry that are used.

I. Cut Locus and of Metric Spheres

We shall start by giving definitions and notations used throughout this article. Let M be a connected, complete and noncompact Riemannian manifold without boundary. For each point $p \in M$, there is an open set U in the tangent space M_p to M at p, which is star shaped with respect to the origin and which is the maximal open set on which the exponential map $\exp_p: M_p \to M$ at p has maximal rank and one to one. The *tangent cut locus* C_p at p is defined to be $C_p = \bar{U} - U$, where $\bar{U}$ is the closure of U. The *cut locus* $C(p)$ at p is defined to be $C(p) = \exp_p(C_p)$. For a fixed point p on M and for a positive number r, let $B_r(p) = \{x \in M; d(p,x) < r\}$, and let $S_r(p) = \{x \in M; d(p,x) = r\}$, where d is by definition the distance function.

It is obvious that for any $x \in S_r(p)$ and for any $r' \in (0,r]$, $d(x, S_{r'}(p)) = r - r'$. However it is not necessarily true that for any $x' \in S_{r'}(p)$. $d(x', S_r(p)) = r - r'$. This is because every minimizing geodesic $\sigma: [0,r'] \to M$ with $\sigma(0) = p$ and $\sigma(r') = x'$ does not have its extension $\sigma | [0,r]$ which is minimizing. For such a point x', we therefore have $d(x', S_r(p)) > r - r'$. Note that $S_r(p)$ is not connected in general. We also note that $S_a(p)$ does not intersect $S_b(q)$ in general even if $|a-b| < d(p,q) < a+b$. This phenomenon is caused by the behavior of cut loci $C(p)$ and $C(q)$ at p and q respectively.

But if we restrict ourselves to consider the special case where $\dim M = 2$ and where the Gaussian curvature G is nonnegative, these assumptions make it possible to give an estimate for upper bound of $\{d(x', S_r(p));\ x' \in S_{r'}(p)\}$ for some r, r' with $r > r'$.

From now let M^2 be a connected, complete and noncompact Riemannian 2-manifold without boundary whose Gaussian curvature G is nonnegative everywhere. In order to avoid the trivial case, we may restrict to consider that G is not identically zero. It turns out that M^2 is diffeomorphic to R^2. This is because by a classification theorem of such M^2 that admits a locally nonconstant convex function (see Theorem E of [7] or [3]), M^2 is diffeomorphic to an plane, a cylinder $S^1 \times R$ or an open Möbius strip. And in the latter cases it turns out that G is identically zero.

Proposition 1.1. *Let $p \in M^2$ be a fixed point and let $a > 0$ be a fixed number. Then there exists a monotone increasing divergent sequence $\{r_j\}$ so that if $b \in [0,a]$ is any and if $x \in S_{r_j+b}(p)$ is any point, then $d(x, S_{r_j+b}(p)) \leqq a+1$.*

Proof. Let $S^1 = \{u \in M_p;\ \|u\| = 1\}$ be the unit circle and let $\Pi: \bar{U} \to S^1$ be the projection. Since $\Pi | C_p$ is one to one, $\Pi(C_p)$ has its measure $m(\Pi(C_p))$ at most 2π. For $r > 0$, let $U_r = \{u \in U;\ \|u\| = r\}$. Since U is star shaped with respect to the origin and $C_p=\bar{U}-U$, $r\mapsto m(\Pi(U_r))$ is monotone deacreasing with $r>0$ and $m(\Pi(U_r))=2\pi$ for small r so that U_r does not intersect C_p. For a fixed $r_0>0$, we have

$$m(\Pi(U_{r0}))-m(\Pi(U_{r0+ka}))=\sum_{i=0}^{k-1}[m(\Pi(U_{r0+ia}))-m(\Pi(U_{r0+(i+1)a}))].$$

Therefore if the measure of $\Pi(U_{r0+ia})-\Pi(U_{r0+(i+1)a})$ is not smaller than $\{r_0+(i+1)a\}^{-1}$ for all $i=0,1,\ldots$, then

$$\lim_{k\to\infty}[m(\Pi(U_{r0}))-m(\Pi(U_{r0+ka}))]\geqq\lim_{k\to\infty}\sum_{l=0}^{k-1}\{r_0+(i+1)a\}^{-1}=\infty.$$

Thus we must have a divergent sequence $\{n_j\}$ of natural numbers with $m(\Pi(U_{r0+n_ja}))-m(\Pi(U_{r0+(n_j+1)a}))<\{r_0+(n_j+1)a\}^{-1}$. Letting $r_r=r_0+n_ja$, we observe that for any $b\in[0,a]$ and for any $u\in U_{r_j}$, there is a $v\in U_{r_j+b}$ with the property that $\measuredangle(u,v)<(r_j+a)^{-1}$.

To estimate the distance between $\exp_p u$ and $\exp_p v$, we construct a piecewise regular curve in $\bar{U}$ which connects u to v and whose interior is in U. First join u to $(r_j/(r_j+b))\,v$ by a circle arc in U with radius r_j and center at the origin. Then connect $(r_j/(r_j+b))\,v$ to v by a straight line segment. By means of Rauch's theorem we see that the exponentiated image of this curve has length at most $a+1$. This proves our proposition.

For a fixed point $p\in M$, we shall define a function δ_p: $[0,\infty)\to M$ by

$$\delta_p(t)=\text{the diameter of } S_t(p),\qquad t>0.$$

Since $S_t(p)$ is not necessarily connected, δ_p is not continuous in general. But we have for any t, t' with $t\geqq t'>0$, $\delta_p(t)-\delta_p(t')\leqq 2(t-t')$. This is indeed if $x,y\in S_t(p)$ are such that $d(x,y)=\delta_p(t)$ and if $x',y'\,S_{t'}(p)$ are intersections of $S_{t'}(p)$ with minimizing geodesics from p to x, y respectively, then $\delta_p(t)=d(x,y)\leqq d(x,x')+d(x',y')+d(y',y)\leqq\delta_p(t)+2(t-t')$. We also note that $0<\delta_p(t)/t\leqq 2$ holds for all $t>0$. Therefore we can find a monotone increasing divergent sequence $\{t_j\}$ so that $\lim_{j\to\infty}(\delta_p(t_j)/t_j)$ exists.

Theorem 1.2. *Let $p\in M^2$ be a fixed point. Then there exists a monotone increasing divergent sequence $\{t_j\}$ so that $\lim_{j\to\infty}(\delta_p(t_j)/t_j$ exists and so that for any point $q\in M^2$ we have*

$$\lim_{j\to\infty}\frac{\delta_p(t_j)}{t_j}\geqq\limsup_{j\to\infty}\frac{\delta_q(t_j)}{t_j}.$$

Proof. Let $\{r_j\}$ be a monotone increasing divergent sequence obtained in Proposition 1.1 for $p\in M^2$ and for $a=d(p,q)$. Set $t_j=r_j+a$. Then for any $b\in[0,a]$ and for any $y\in S_{t_j-b}(p)$, we have $d(y,S_{t_j}(p))\leqq a+1$.

We shall assert that if $x\in S_{t_j}(q)$ is any point, then $d(x,S_{t_j}(p))\leqq a+1$. To show this we first take a point $x\in S_{t_j}(q)-B_{t_j}(p)$. Then $d(x,S_{t_j}(p))=a$, because for a minimizing geodesic σ: $[0,l]\to M^2$ with $p=\sigma(0)$ and $x=\sigma(l)$ we have $\sigma(t_j)\in S_{t_j}(p)$. From triangle inequality we see $l\leqq t_j+a$, and hence $d(x,S_{t_j}(q))\leqq l-t_j\leqq a$. Next, we take

$x \in S_{t_j}(q) \cap B_{t_j}(p)$ and let $r' = d(p, x)$, and $b = t_j - r'$. We then see that $b = t_j - r' \leqq a$. Since $x \in S_{t_j - b}(p)$, Proposition 1.1 implies the desired inequality.

Now let us take for each j, a pair of points $x_{q,j}, y_{q,j} \in S_{t_j}(q)$ with $d(x_{q,j}, y_{q,j}) = \delta_q(t_j)$. Then we can find $x_{p,j}$ and $y_{p,j}$ on $S_{t_j}(p)$ so that $d(x_{p,j}, x_{q,j}) \leqq a+1$ and $d(y_{p,j}, y_{q,j}) \leqq a+1$. Therefore we have $\delta_p(t_j) \geqq d(x_{p,j}, y_{p,j}) \geqq \delta_q(t_j) - 2(a+1)$ for all j. This completes the proof.

For each point $x \in M$, the *injectivity radius* $i(x)$ at x is defined by $i(x) = \inf\{\|u\|; u \in C_x\}$. It is well known that $i: M \to R$ is continuous.

Lemma 1.3. *Let $\gamma: [0, \infty) \to M$ be a ray. If there is a monotone increasing divergent sequence $\{t_j\}$ so that $i(\gamma(t_j)) \geqq t_j$, then $F_\gamma: M \to R$ is not exhaustion.*

Proof. It suffices to show that $F_\gamma^{-1}(\{0\})$ is not compact. Note that $p = \gamma(0)$ belongs to $F_\gamma^{-1}(\{0\})$. For any positive R, we want to find a point x on $F_\gamma^{-1}(\{0\})$ with $d(x, p) = R$. For this purpose we fix j_0 so that $t_{j_0} > R$ and let $K_0 = S_R(p) - B_{t_{j_0}}(\gamma(t_{j_0}))$. The assumption $i(\gamma(t_j)) \geqq t_j$ ensures connectness of $S_{t_j}(\gamma(t_j))$. Note that if $j > j_0$, then p and $\gamma(2t_j)$ are on $S_{t_j}(\gamma(t_j))$ and $d(p, \gamma(2t_j)) = 2t_j > R$. Therefore we can find an $x_j \in S_{t_j}(\gamma(t_j)) \cap S_R(p)$ from continuity of distance function to p which is restricted on $S_{t_j}(\gamma(t_j))$. We also see that if $j > j_0$, then $B_{t_{j_0}}(\gamma(t_{j_0})) \subset B_{t_j}(\gamma(t_j))$, and hence $x_j \in K_0$. Since K_0 is compact we can choose a subsequence of $\{x_j\}$ which tends to a point $x \in K_0$. It follows from the definition of F_γ that x belongs to $F_\gamma^{-1}(\{0\})$. Because $d(x, p) = R$ and R is any, F_γ is not exhaustion. This completes the proof.

II. Non-Exhaustion Busemann Functions

In the previous section the assumption that $\dim M = 2$ was essential. But since the essential tool used here is the triangle comparison theorem, we can deal with general case where dimension of M is $m \geqq 2$ on which the diameter $\delta_p(t)$ of metric t-sphere $S_t(p)$ at p is naturally defined in the same way as in § 1. Throughout this section let M be an m-dimensional connected, complete and noncompact Riemannian manifold whose sectional curvature is nonnegative everywhere.

Proposition 2.1. *Assume that there is a point p and there is a monotone increasing divergent sequence $\{t_j\}$ with $\delta_p(t_j)/t_j > \sqrt{2}$ for all j. Then there exist rays σ, τ: $[0, \infty) \to M$ emanating from $p = \sigma(0) = \tau(0)$ with $\langle \dot\sigma(0), \dot\tau(0) \rangle \leqq 0$ and with respect to which the Busemann functions F_σ and F_τ are non-exhaustion.*

Proof. Let $x_j, y_j \in S_{t_j}(p)$ be such that $\delta_p(t_j) = d(x_j, y_j)$ and let σ_j, τ_j: $[0, t_j] \to M$ be minimizing geodesics with $\sigma_j(0) = \tau_j(0) = p$ and $\sigma_j(t_j) = x_j$, $\tau_j(t_j) = y_j$. We then observe that $\sigma_j([0, t_j]) \subset M - B_{t_j}(y_j)$ and $\tau_j([0, t_j]) \subset M - B_{t_j}(x_j)$. In fact suppose that $\sigma_j([0, t_j])$ intersects $B_{t_j}(y_j)$. Then there is a point q_j on $\sigma_j((0, t_j])$ so that $d(y_j, q_j) = d(y_j, \sigma_j([0, t_j]) < t_j$. Since a minimizing geodesic joining q_j to y_j is orthogonal to σ_j at q_j, and since $d(y_j, q_j) < t_j$, $d(x_j, q_j) < t_j$ and $d(x_j, y_j) > \sqrt{2} t_j$, we derive a contradiction by applying the triangle comparison theorem to the triangle with vertices (y_j, q_j, x_j).

Now let σ, τ: $[0, \infty) \to M$ be the rays obtained by $\dot\sigma(0) = \lim_{j \to \infty} \dot\sigma_j(0)$ and $\dot\tau(0) = \lim_{j \to \infty} \dot\tau_j(0)$. From the triangle comparison theorem we see that $\langle \dot\sigma(0), \dot\tau(0) \rangle \leqq 0$. We also note that every Busemann function on M is convex because the sectional curvature of M is nonnegative everywhere.

Finally we shall assert that $F_\sigma: M\to R$ is not exhaustion. This is achieved by showing that $F_\sigma\circ\tau: [0,\infty)\to R$ is monotone non-increasing. In fact we know from Theorem B of [7] that if a locally nonconstant convex function has a compact level, then all of level sets are compact. Therefore if F_σ takes minimum and if $F_\sigma\circ\tau$ is monotone non-increasing, then there is a noncompact level of F_σ, and thus F_σ is non-exhaustion. If F_σ does not take minimum, then clearly F_σ is non-exhaustion.

The basic idea of proving that $F_\sigma\circ\tau: [0,\infty)\to R$ is monotone non-increasing is due to Alexandrov's convexity theorem (see Lemma 19, p. 318 of [13]), which is stated as follows: Let α,β: $[0,1]\to M$ be minimizing geodesics parametrized proportionally to arc length on $[0,1]$ with $\alpha(0)=\beta(0)$. For each $s,t\in[0,1]$ let $\Delta(s,t)$ be the triangle sketched on R^2 whose edge lengths are $s\|\dot\alpha\|$, $t\|\dot\beta\|$ and $d(\alpha(s),\beta(t))$, (where $\dot\alpha$ means the velocity vector of α and $\|\dot\alpha\|$ its norm). Let $\theta(s,t)$ be the edge angle of $\Delta(s,t)$ opposite to the edge of length $d(\alpha(s),\beta(t))$. Then $(s,t)\mapsto\theta(s,t)$ is monotone non-increasing in the following sense: If $s_1,s_2,t_1,t_2\in[0,1]$ are taken so that $s_1\leqq s_2$ and $t_1\leqq t_2$, then $\theta(s_1,t_1)\geqq\theta(s_2,t_2)$.

To each geodesic triangle with vertices $(p,\sigma_j(t_j),\tau_j(t_j))$ we apply Alexandrov's convexity theorem and making use of $\delta_p(t_j)=d(\sigma_j(t_j),\tau_j(t_j))>\sqrt{2}t_j$, we obtain $d(\tau_j(u),\sigma_j(v))-v>u/((u^2+v^2)^{1/2}+v)$ for all $u,v\in[0,t_j]$. Since $\sigma(v)=\lim_{j\to\infty}\sigma_j(v)$, we find a large j' so that if $j>j'$, then

$$\begin{aligned}d(\sigma(v),\tau_j(u))-v&\geqq d(\sigma_j(v),\tau_j(u))-d(\sigma_j(v),\sigma(v))-v\\&>u/2\{(u^2+v^2)^{1/2}+v\}>0,\end{aligned}$$

and hence $\tau_j(u)\notin B_v(\sigma(v))$ for all $j>j'$. Because u and v are any, we have $\tau_j(u)\notin\bigcup_{v>0}B_v(\sigma(v))$. Therefore from $\tau(u)=\lim_{j\to\infty}\tau_j(u)$, follows $\tau(u)\notin\bigcup_{v>0}B_v(\sigma(v))$. Since

$$F_\sigma^{-1}(\{0\})=\overline{\bigcup_{v>0}B_v(\sigma(v))}-\bigcup_{v>0}B_v(\sigma(v)),$$

we see $F_\sigma\circ\tau(u)\leqq 0$.

Proposition 2.2. *Let α: $[0,\infty)\to M$ be a ray emanating from p. Assume that $F_\alpha: M\to R$ is not exhaustion and assume that F_α does not take minimum. Then $\delta_p(t)/t>\sqrt{2}$ for all $t>0$.*

Proof. Since $F_\alpha^{-1}((-\infty,0])=M-\bigcup_{t>0}B_t(\alpha(t))$ is noncompact, closed and totally convex, and since p is on it, there is a ray β: $[0,\infty)\to F_\alpha^{-1}((-\infty,0])$ on which $F_\alpha\circ\beta$: $[0,\infty)\to\mathbb{R}$ is a strictly monotone decreasing convex function. Such a β can be obtained as follows: Let $\{x_j\}$ be a divergent sequence of points in $F_\alpha^{-1}((-\infty,0])$ such that $F_\alpha(x_j)$ is strictly decreasing with j and $\lim_{j\to\infty}F_\alpha(x_j)=\inf F_\alpha$. Let β_j be a minimizing geodesic joining $p=\beta_j(0)$ to x_j with $\|\dot\beta_j\|=1$. Then β is obtained by the limit $\dot\beta(0)=\lim_{j\to\infty}\dot\beta_j(0)$. It is now clear from construction that $F_\alpha\circ\beta$ is strictly monotone decreasing.

For every $t>0$ let α_t: $[0,\infty)\to M$ be a ray emanating from $\beta(t)=\alpha_t(0)$ which is *asymptotic* to α. Namely, α_t is defined as follows: $\dot\alpha_t(0)$ is the limit of unit vectors $\{\dot\alpha_{i,j}(0)\}$ where $\alpha_{i,j}$: $[0,l_j]\to M$ is a minimizing geodesic with $\alpha_{i,j}(0)=\beta(t)$ and $\{\alpha_{i,j}(l_j)\}$ is a divergent sequence on $\alpha([0,\infty))$. Letting $-a_t=F_\alpha(\beta(t))$, we observe that $\dot\alpha_t(0)$ is transversal to

$$F_\alpha^{-1}(\{-a_t\})=\overline{\bigcup_{s>0} B_{a_t+s}(\alpha(s))}-\bigcup_{s>0} B_{a_t+s}(\alpha(s))$$

and $\alpha_t([0,\infty))$ is a radial to $\bigcup_{s>0} B_{a_t+s}(\alpha(s))$ through $\beta(t)$. We then find an $\varepsilon_t>0$ so that $\sphericalangle(-\dot\beta(t),\dot\alpha_t(0))\leqq\pi/2-\varepsilon_t$. Indeed we first note that $F_\alpha(\alpha_t(s))=s-a_t$ and F_α is differentiable in a neighborhood of $\alpha_t((0,\infty))$. Obviously $\nabla F_\alpha(\alpha_t(s))=\dot\alpha_t(s)$ for all $s>0$ (, where $\nabla F_\alpha(x)$ is by definition the gradient vector for F_α at x). For small $s>0$ so that $\alpha_t(s)\notin C(p)$ and F_α is differentiable at $\alpha_t(s)$, there is a unique shortest connection β_s: $[0,t_s]\to M$ with $\beta_s(0)=p$, $\beta_s(t_s)=\alpha_t(s)$ and $t_s=d(p,\alpha_t(s))$. Then

$$\begin{aligned}\langle\nabla F_\alpha,-\dot\beta_s(t_s)\rangle&=-\dot\beta_s(t_s)(F_\alpha)=\lim_{h\downarrow 0}\{F_\alpha\circ\beta_s(t_s-h)-F_\alpha\circ\beta_s(t_s)\}/h\\&\geqq\{F_\alpha\circ\beta_s(0)-F_\alpha\circ\beta_s(t_s)\}/t_s=(a_t-s)/t_s.\end{aligned}$$

Let $\varepsilon_t=\sin^{-1}a_t/t_s$.

We shall finally apply Alexandrov's convexity theorem to see that if $\theta(t,s)$ is the edge angle of the triangle $\Delta(s,t)$ sketched on R^2 with edge length s, t and $d(\alpha(s),\beta(t))$ and it is opposite to the edge of length $d(\alpha(s),\beta(t))$, then $s\mapsto\theta(s,t)$ is monotone nonincreasing for every fixed $t>0$. Moreover we have $\lim_{s\to\infty}\theta(s,t)=\pi$ minus limit of the edge angle of $\theta(s,t)$ opposite to the edge of length $s\geqq\pi-\sphericalangle(\dot\alpha_t(0),-\dot\beta(t))\geqq\pi/2+\varepsilon_t$. Therefore $\theta(s,t)\geqq\lim_{s\to\infty}\theta(s,t)\geqq\pi/2+\varepsilon_t$ implies that $\theta(t,t)\geqq\pi/2+\varepsilon_t$, and hence $d(\alpha(t),\beta(t))>\sqrt{2}t$ for all $t>0$. Since $\delta_p(t)=d(\alpha(t),\beta(t))$, this completes the proof.

Proposition 2.3. *Let γ: $[0,\infty)\to M$ be a ray and let $x,y\in F_\gamma^{-1}(\{a\})$ for a fixed a. Assume that there exists a minimizing geodesic α: $[0,l]\to M$ with $\alpha(0)=x$, $\alpha(l)=y$, $l=d(x,y)$ and a ray σ: $[0,\infty)\to M$ which is asymptotic to γ with $\sigma(0)=x$ so that $\langle\dot\alpha(0),\dot\sigma(0)\rangle=0$. Then there exists a flat totally geodesic surface whose boundary consists of $\alpha([0,l])$, $\sigma([0,\infty))$ and $\tau([0,\infty))$, and τ is the ray with $\tau(0)=y$ and asymptotic to γ and $\dot\tau(0)$ is parallel to $\dot\sigma(0)$ along α.*

Proof. From definition of asymptotic ray, we have a divergent sequence $\{t_j\}$ and a family $\{\sigma_j$: $[0,l_j]\to M, j=1,2,\dots\}$ of minimizing geodesics so that $\sigma_j(0)=x$, $\sigma_j(l_j)=\gamma(t_j)$, $d(x,\gamma(t_j))=l_j$ for all j and $\dot\sigma(0)=\lim_{j\to\infty}\dot\sigma_j(0)$. For each $s\in[0,l]$ and for each j, let $\sigma_{s,j}$: $[0,u_{s,j}]\to M$ be a minimizing geodesic from $\alpha(s)=\sigma_{s,j}(0)$ to $\gamma(t_j)=\sigma_{s,j}(u_{s,j})$. From convexity of F_γ follows $F_\gamma\circ\alpha(s)\leqq a$ for all $s\in[0,l]$ and hence $\lim_{j\to\infty}[u_{s,j}-l_j]\geqq 0$, especially we have $\lim_{j\to\infty}[u_{l,j}-l_j]=0$. For every $s\in[0,l]$ we apply triangle comparison theorem to the triangle $(x,\alpha(s),\gamma(t_j))$, and letting $j\to\infty$, we obtain from $\langle\dot\alpha(0),\dot\sigma(0)\rangle=0$ that $\lim_{j\to\infty}[u_{s,j}-l_j]=0$. This means that if σ_s: $[0,\infty)\to M$ is a ray emanating from $\alpha(s)$ and asymptotic to γ so that $\dot\sigma_s(0)=\lim_{j\to\infty}\dot\sigma_{s,j}(0)$, then $\langle\dot\sigma_s(0),\dot\alpha(s)\rangle$

$=0$, and therefore $\alpha([0,l])\in F_\gamma^{-1}(\{a\})$. Note that if $\tau\colon[0,\infty)\to M$ is a ray asymptotic to γ, then $F_\gamma\circ\tau(u)=u+F_\gamma\circ\tau(0)$ for all $u\geqq 0$.

Let $b>a$ be arbitrary fixed. From Theorem 1.10 of [3], $s\mapsto -d(\alpha(s),F_\gamma^{-1}(\{b\}))$ is convex for $s\in[0,l]$. For every $s\in[0,l]$, let $\mathbf{P}_s$ be the unit parallel field along $\sigma_s|[0,b-a]$ with the initial condition $\mathbf{P}_s(0)=\dot\alpha(s)$. It follows from total convexity of $F_\gamma^{-1}((-\infty,b])$ that for every $s\in[0,l]$ the 1-parameter variation defined by $V_s(\varepsilon,t)=\exp_{\sigma_s(t)}\varepsilon\mathbf{P}_s(t)$ has the property that $F_\gamma\circ V_s(\varepsilon,b-a)\geqq b$. But because the sectional curvature is nonnegative everywhere, the second variation for this V_s is nonpositive. Thus we must have that the variation curve $t\mapsto V_s(\varepsilon,t)$ is a geodesic and $V_s(\varepsilon,b-a)\in F_\gamma^{-1}(\{b\})$ for every ε. Thus the map $(\varepsilon,t)\mapsto V_s(\varepsilon,t)$ defines a piece of flat totally geodesic surface for every $s\in[0,l]$, and $\varepsilon\mapsto V_s(\varepsilon,b-a)$ is a geodesic which is contained in $F_\gamma^{-1}(\{b\})$.

We start this construction from $s=0$ and $V_0(\varepsilon,t)=\exp_{\sigma(t)}\varepsilon\mathbf{P}_0(t)$. By repeating this process we obtain a piece of flat rectangle totally geodesic surface whose boundary consists of $\alpha([0,l])$, $\sigma([0,b-a])$ and their opposite sides. For every $s\in(0,l]$ let $\tilde\sigma_s\colon[0,b-a]\to M$ be the minimizing geodesic on the surface so that $\dot{\tilde\sigma}_s(0)$ is parallel to $\dot\sigma(0)$ along α and $\tilde\sigma_s(0)=\alpha(s)$. It follows from the definition of Busemann function that the extension $\tilde\sigma_s\colon[0,\infty)\to M$ becomes a ray which is asymptotic to σ. Because $b>a$ is any and the curve $s\mapsto\tilde\sigma_s(b-a)$ is a geodesic in $F_\gamma^{-1}(\{b\})$, we can extend the flat totally geodesic surface whose boundary consists of $\sigma([0,\infty))$, $\alpha([0,l])$ and $\tau([0,\infty))$. This completes the proof.

With the aids of the above Propositions we have the following

Theorem 2.4. *Assume that there is a point $x\in M$ and that there is a monotone increasing divergent sequence $\{t_j\}$ so that $\delta_x(t_j)/t_j>\sqrt{2}$ holds for all j. Then we have:*

(1) *$\delta_x(t)/t\geqq\sqrt{2}$ for all $t>0$.*

(2) *For every point $y\in M$, $\delta_y(t)/t\geqq\sqrt{2}$ for all $t>0$.*

(3) *There are at least two rays emanating from every point $y\in M$, with respect to which the Busemann functions are non-exhaustion.*

Proof. It follows from Proposition 2.1 that there are two rays $\sigma,\tau\colon[0,\infty)\to M$ emanating from x with respect to which the Busemann functions are non-exhaustion, and moreover $F_\sigma\circ\tau\colon[0,\infty)\to R$ is, as is shown in the proof of Proposition 2.2, monotone nonincreasing so that $\lim_{t\to\infty}F_\sigma\circ\tau(t)=\inf F_\sigma$.

Now, if $F_\sigma\circ\tau|[t^*,\infty)$ is constant for some $t^*\geqq 0$, then $\langle\dot\sigma_t(0),\dot\tau(t)\rangle=0$ holds for every $t\geqq t^*$ and for every ray $\sigma_t\colon[0,\infty)\to M$ with $\sigma_t(0)=\tau(t)$ and which is asymptotic to σ. It follows from Proposition 2.3 that there is a flat totally geodesic surface with boundary $\tau([t^*,\infty))$ and $\sigma_{t^*}([0,\infty))$ where σ_{t^*} is a ray emanating from $\tau(t^*)$ and asymptotic to σ. If there is a $t'\in[0,t^*)$ with $F_\sigma\circ\tau(t')>F_\sigma\circ\tau(t^*)$, then $\langle\dot\sigma(0),\dot\tau(0)\rangle<0$. In this case we have $d(\sigma(t),\tau(t))>\sqrt{2}t$ for all $t>0$. If $t^*=0$, then $d(\sigma(t),\tau(t))=\sqrt{2}t\leqq\delta_p(t)$ for all $t>0$.

In the case where $F_\sigma\circ\tau$ is strictly decreasing, we see that $\langle\dot\sigma(0),\dot\tau(0)\rangle<0$. Therefore $\delta_p(t)\geqq d(\sigma(t),\tau(t))>\sqrt{2}t$ holds for all $t>0$.

To prove (2) and (3), let $\sigma_y,\tau_y\colon[0,\infty)\to M$ be rays emanating from $y=\sigma_y(0)=\tau_y(0)$ and asymptotic to σ,τ respectively. We note that both $F_\sigma\circ\tau_y$ and $F_\tau\circ\sigma_y$ are monotone nonincreasing convex functions. Let $F_\sigma(y)=a_0$ and $F_\tau(y)=b_0$. It follows from a slight extension of Lemma 2.3 of [12] that $F_\tau^{-1}((-\infty,b_0])$

$\subset F_{\tau_y}^{-1}((-\infty,0])$ and $F_\sigma^{-1}((-\infty,a_0])\subset F_{\sigma_y}^{-1}((-\infty,0])$. This fact means that both $F_{\sigma_y}\circ\tau_y$ and $F_{\tau_y}\circ\sigma_y$ are nonpositive convex functions defined on $[0,\infty)$. Therefore F_{σ_y} and F_{τ_y} are non-exhaustion. From the above argument we obtain $d(\sigma_y(t),\tau_y(t))\geqq\sqrt{2}t$ for all $t>0$. Thus (2) and (3) are proved.

III. Non-Exhaustion Busemann Functions and Total Curvature

In this and the next section we shall deal with connected, noncompact, complete Riemannian 2-manifold M^2 which has no boundary and whose Gaussian curvature G is nonnegative everywhere. Here we want to prove that if M^2 admits a non-exhaustion Busemann function, then the total curvature is not greater than π. As is stated in § 1, we may consider that G is not identically zero, thus M^2 is diffeomorphic to R^2.

Theorem 3.1. *Let γ: $[0,\infty)\to M^2$ be a ray with respect to which F_γ is non-exhaustion. If F_γ takes minimum, then we have the following:*

(1) *$\delta_x(t)/t>\sqrt{2}$ for all $t>0$ and for all $x\in M^2$.*

(2) *The support of G is contained in a closure compact domain bounded by a geodesic biangle with the edge angles $\pi/2$ if and only if the Busemann function with respect to the minimum set of F_γ takes minimum.*

(3) *The total curvature of M^2 is not greater than π.*

Remark. It follows from Theorem B of [7] that the minimum set of F_γ is non-compact, which is either a ray or a straight line. If the minimum set is a straight line, then the splitting theorem (see [14]) implies that G is identically zero. Furthermore a theorem due to Cohn-Vossen (see Satz 5 in [6]) states: Let $\mathscr{F}$ be a complete open 2-manifold diffeomorphic to R^2. If $\mathscr{F}$ contains a straight line and if the total curvature of $\mathscr{F}$ exists, then the total curvature is nonpositive. Thus in our case the minimum set is a ray.

Proof. Let σ: $[0,\infty)\to M^2$ be the ray whose image is the minimum set of F_γ. We shall first of all assert that through each point $\sigma(t)$, $t\geqq 0$, there passes exactly two rays α_t^+ and α_t^- with $\langle\dot\alpha_t^+(0),\dot\sigma(t)\rangle=\langle\dot\alpha_t^-(0),\dot\sigma(t)\rangle=0$, and they are asymptotic to γ. This is in fact there is at least a ray emanating from $\sigma(t)$ and asymptotic to γ, say, α_t^+: $[0,\infty)\to M^2$. Take $t'>t$ and apply triangle comparison theorem to the geodesic triangle with vertices $\sigma(t)$, $\sigma(t')$ and $\gamma(s)$. Letting $s\to\infty$, we get $\langle\dot\alpha_t^+(0),\dot\sigma(t)\rangle=0$. Then for each $\varepsilon>0$, there is an $\eta>0$ with $\lim_{s\to\infty}[d(\alpha_t^+(-\varepsilon),\gamma(s))-d(\sigma(t),\gamma(s))]=-\eta$, and hence every ray emanating from $\alpha_t^+(-\varepsilon)$ and asymptotic to γ makes an angle with $\dot\alpha_t^+(-\varepsilon)>\pi/2$. Taking the limit of wuch asymptotic rays by letting $\varepsilon\to 0$, we find another ray α_t^-: $[0,\infty)\to M^2$ which is emanating from $\sigma(t)$ and asymptotic to γ and $\langle\dot\alpha_t^-(0),\dot\sigma(t)\rangle=0$.

Because there are exactly two rays emanating from each $\sigma(t)$ and asymototic to γ, we may choose α_t^+ so that $t\to\dot\alpha_t^+(0)$ is parallel along σ. From Proposition 2.3 we see that there is an isometric embedding I: $R\times[0,\infty)\to M^2$ with $I(s,t)=\alpha_t^+(s)$ for $s\in R$ and $t\in[0,\infty)$, and $I(0,t)=\sigma(t)$ for all $t\geqq 0$. $\mathscr{S}=I(R\times[0,\infty))$ is a flat totally geodesic surface.

Next, we shall assert that

$$\int_{M^2-\mathscr{S}} G\,d\mu \leqq \pi.$$

For simplicity we set $\alpha_t=\alpha_t^+$. Let $-a=\min F_\gamma$. Since $\mathscr{S}$ is isometric to a closed half space of R^2 and since $\dot\sigma(t)$ is orthogonal to $\dot\alpha_t(0)$ for all $t\geqq 0$, we see that $F_\sigma^{-1}(\{t\})=\alpha_t(R)$ for all $t\geqq 0$. $F_\sigma((-\infty,0])=M^2-\mathscr{S}$ is a closed totally convex set in which $\gamma([0,\infty))$ lies. Thus F_σ is non-exhaustion and $F_\sigma\circ\gamma$ is a monotone nonincreasing convex function. For each $t>0$ let λ_t: $[0,l_t]\to M^2$ be a minimizing geodesic with $\lambda_t(0)=\alpha_0(t)$, $\lambda_t(l_t)=\alpha_0(-t)$. There is a $t_1>0$ so that if $t>t_1$, then $d(\alpha_0(-t),\alpha_0(t))<2t$. If $t>t_1$ then λ_t does not coincide with $\alpha_0|[-t,t]$. By virtue of convexity of $F_\gamma\circ\lambda_t$: $[0,l_t]\to R$ which takes value $t-a$ at both ends and because of $\nabla F_\gamma\circ\alpha_0(\pm t)=\pm\dot\alpha_0(\pm t)$ for all $t>0$, λ_t satisfies that $\measuredangle(\dot\lambda_t(0),\dot\alpha_0(t))\geqq\pi/2$ and $\measuredangle(\dot\lambda_t(l_t),\ \dot\alpha_0(-t))\geqq\pi/2$. Let D_t be the domain bounded by $\lambda_t([0,l_t])$ and $\alpha_0([-t,t])$ which does not intersect $\sigma([0,\infty))$. Then $p\in D_t$ and we see from Gauss-Bonnet theorem that

$$\int_{D_t} G\,d\mu+\measuredangle(\dot\lambda_t(0),\dot\alpha_0(t))+\measuredangle(\dot\lambda_t(l_t),\ \dot\alpha_0(-t))=2\pi.$$

We see from Lemma 2.2 of [12] that $t\mapsto l_t$ is monotone increasing for $t\geqq 0$ and it is constant on $[t',\infty)$ if and only if there exists a flat totally geodesic surface with boundary $\alpha_0([t',\infty))$, $\alpha_0((-\infty,-t'])$ and $\lambda_{t'}([0,l_{t'}])$. In this case the support of G is contained in $D_{t'}$ which is bounded by the geodesic biangle with sides $\alpha_0([-t',t'])$ and $\lambda_{t'}([0,l_{t'}])$ and with edge angles $\pi/2$. Therefore we have in this case

$$\int_{M^2} G\,d\mu=\int_{D_{t'}} G\,d\mu=\pi.$$

We also observe in this case that F_σ takes minimum which is equal to $-l_{t'}/2$. Conversely, if F_σ takes minimum, then the minimum set of F_σ is a ray $\tilde\gamma$: $[0,\infty)\to M^2$. Because $d(\tilde\gamma(t),\alpha_0(R))$ is equal to $-\min F_\sigma$, we find a flat totally geodesic embedding I': $R\times[0,\infty)\to M^2$ with $I'(0,t)=\tilde\gamma(t)$ and $s\mapsto I'(s,t)$ forms geodesic which is orthogonal to $\tilde\gamma(t)$ at the origin for every $t\geqq 0$. Let $\mathscr{S}'=I'(R\times[0,\infty))$. Then the support of G is contained in $M^2-\mathscr{S}\cup\mathscr{S}'$ which is a domain bounded by a geodesic biangle with edge angles $\pi/2$, and which does not intersect $\sigma([0,\infty))$ and $\tilde\gamma([0,\infty))$. This proves (2).

Thus we may consider the case where one of the two edge angles of every geodesic biangle which bounds D_t is less than $\pi/2$. In this case we have

$$\int_{D_t} G\,d\mu<\pi$$

for all $t\geqq t_1$, where t_1 is taken to be $2t_1>d(\alpha_0(-t_1),\alpha_0(t_1))$. Therefore we get from $\bigcup_{t>t_1} D_t\cup\mathscr{S}=M^2$ and $\{D_t\}$ being monotone increasing that

$$\int_{M^2} G\,d\mu\leqq\pi.$$

These two cases complete the proof of (3).

We shall finally prove (1). If F_σ does not take minimum and if $x \in M^2$ is any point, then there exist rays σ_x, τ_x: $[0, \infty) \to M^2$ with $\sigma_x(0) = \tau_x(0) = x$ so that σ_x is asymptotic to σ and so that $\lim_{t \to \infty} F_\sigma \circ \tau_x(t) = \inf F_\sigma$. Since $F_{\sigma_x} \leqq F_\sigma - F_\sigma(x)$ follows from a slight extension of Lemma 2.3 in [12], we see $F_{\sigma_x} \circ \tau_x$: $[0, \infty) \to R$ is strictly monotone decreasing. Therefore we have from Proposition 2.2 that $\delta_x(t)/t > \sqrt{2}$ for all $t > 0$.

Thus in order to prove (1), we may restrict to consider the case where F_σ takes minimum. Let $-a = \min F_{\tilde{\gamma}}$ and $-b = \min F_\sigma$, where $\tilde{\gamma}$: $[0, \infty) \to M^2$ is the ray whose image is the minimum set of F_σ. It follows from $F_\sigma \circ \tilde{\gamma}$ being constant that $F_{\tilde{\gamma}}$ also takes minimum and the minimum set is $\sigma([0, \infty))$. For simplicity we shall rewrite $\gamma = \tilde{\gamma}$. For every $s \geqq 0$, let β_s: $R \to M^2$ be the geodesic with $\beta_s(0) = \gamma(s)$ so that $\dot{\beta}_s(0)$ is unit parallel field along γ which is orthogonal to $\dot{\gamma}$ and so that $\beta_0(b) = \alpha_0(a)$, $\beta_0(-b) = \alpha_0(-a)$. Then the support of G is contained in D which is bounded by $\alpha_0([-a, a])$ and $\beta_0([-b, b])$. Now if $x \notin D$, then x belongs to either $\mathscr{S}$ or else $\mathscr{S}'$. For each $t > 0$, we can find points y, z on $\mathscr{S} \cup \mathscr{S}'$ so that (x, y, z) forms an isosceles triangle with $d(x, y) = d(x, z) = t$ and $d(y, z) > \sqrt{2}t$. If $x \in D$, then we take rays σ_x, γ_x: $[0, \infty) \to M^2$ with $\sigma_x(0) = \gamma_x(0) = x$ and so that they are asymptotic to σ and γ respectively. Setting $-a' = F_\gamma(x)$ and $-b' = F_\sigma(x)$, we see that $F_\gamma \circ \sigma_x$: $[0, \infty) \to R$ is monotone nonincreasing and it is constant $= -a''$ on $[b', \infty)$. This is because $\sigma_x(b') = \alpha_0(a - a'')$ and $b' = d(x, \alpha_0(R))$, and $\dot{\sigma}_x(b')$ is parallel to $\dot{\sigma}(0)$ along $\alpha_0 | [0, a - a'']$. Similarly $F_\sigma \circ \gamma_x$: $[0, \infty) \to R$ is monotone nonincreasing and it is constant $= -b''$ on $[a', \infty)$. We note that $\beta_0(b - b'') = \gamma_x(a')$ and $\alpha_0(a - a'') = \sigma_x(b')$. Moreover $a' < a''$ and $b' < b''$ provided if G is positive in the geodesic quadrangle with sides on $\gamma_x, \sigma_x, \alpha_0$ and β_0. Otherwise G is identically zero in the quadrangle and the previous argument derives the conclusion. If follows from $a' < a''$ and $b' < b''$ that if $t > a'' + b''$, then

$$\begin{aligned}\delta_x(t) &\geqq d(\sigma_x(t), \gamma_x(t)) > \min[\sqrt{2}\{t - (b' + a'')\} + \sqrt{2}(a'' + b''), \sqrt{2}\{t - (a' + b'')\} \\ &\quad + \sqrt{2}(a'' + b'')\}] = \sqrt{2}t + \min\{\sqrt{2}(b'' - b'), \sqrt{2}(a'' - a')\} > \sqrt{2}t.\end{aligned}$$

Thus we see for all $t > 0$ that $\delta_x(t)/t > \sqrt{2}$.

Theorem 3.2. *Let γ: $[0, \infty) \to M^2$ be a ray with respect to which F_γ is non-exhaustion. Then we have*

$$\int_{M^2} G \, d\mu \leqq \pi.$$

Before we go into the proof, we shall first note that if there is a compact level for F_γ, then all levels are compact and hence M^2 has two ends. In this case M^2 is isometric to a flat cylinder and the conclusion is obvious. We may therefore restrict to consider the case where all levels of F_γ are connected and noncompact, and hence they are homeomorphic to R. This is because the existence of a nonconnected level of a locally nonconstant convex function implies that the function takes minimum (see Theorem A of [7]), and this case has already been shown in Theorem 3.1. We may also assume that G is not identically zero and that M^2 does not contain straight line.

Let $p=\gamma(0)$. Then there is a $t_1>0$ so that $\gamma(-t_1)\in C(\gamma(t_1))$. Thus $d(\gamma(-t), \gamma(t))<2t$ holds for all $t>t_1$.

For every $a\leqq 0$ and for every point $x\in F_\gamma^{-1}(\{a\})$, every minimizing geodesic γ_x: $[a,0]\to M^2$ with $\gamma_x(a)=x$, $\gamma_x(0)\in F_\gamma^{-1}(\{0\})$ and $-a=d(x,F_\gamma^{-1}(\{a\}))$ has its extension γ_x: $[a,\infty)\to M^2$ which is a ray emanating from x and asymptotic to γ, (see p. 136, §22 of [2]). Especially if a is taken so that $\inf F_\gamma<a\leqq -t_1$, then for every point $x\in F_\gamma^{-1}(\{a\})$ and for every minimizing geodesic γ_x: $[a,0]\to M^2$ with $\gamma_x(a)=x$, $\gamma_x(0)\in F_\gamma^{-1}(\{0\})$ and $d(x,F_\gamma^{-1}(\{a\}))=-a$, we have $\gamma_x(0)\neq p$. In fact, suppose that there is such a γ_x that satisfies $\gamma_x(0)=p$. Then its extension γ_x: $[a,\infty)\to M^2$ is a ray which obviously contains γ and $\gamma_x(t)=\gamma(t)$ for all $t\geqq 0$, (see §22, Theorem 22.19 of [2]). Then $\gamma(-t_1)\notin C(\gamma(t_1))$, contradicting the choice of t_1.

The above argument ensures the existence of a point $x\in F_\gamma^{-1}(\{a\})$ for every $a\leqq -t_1$ so that there are two minimizing geodesics γ_x,σ_x: $[a,0]\to M^2$ with $\gamma_x(a)=\sigma_x(a)=x$ and $\gamma_x(0)$, $\sigma_x(0)$ are on $F_\gamma^{-1}(\{0\})$ and the two points cut off a subarc of $F_\gamma^{-1}(\{0\})$ which contains p in its interior. Indeed let us denote by A_0 and B_0 the connected and closed subarcs of $F_\gamma^{-1}(\{0\})$ with $A_0\cap B_0=p$, $A_0\cup B_0=F_\gamma^{-1}(\{0\})$ and $\mathring{A}_0\cap B_0=A_0\cap\mathring{B}_0=\emptyset$. For $a\leqq t_1$, let $A_a=\{x\in F_\gamma^{-1}(\{\alpha\})$; if γ_x: $[a,0]\to M^2$ is any minimizing geodesic with $\gamma_x(a)=x$, $\gamma_x(0)\in F_\gamma^{-1}(\{0\})$ and $-a=d(x,F_\gamma^{-1}(\{0\}))$, then $\gamma_x(0)\in A_0\}$. We define $B_a\subset F_\gamma^{-1}(\{a\})$ in the same way as A_a. Suppose no point on $F_\gamma^{-1}(\{a\})$ satisfies the property stated above. Then $F_\gamma^{-1}(\{a\})$ is the disjoint union of A_a and B_a. Obviously A_a and B_a are closed in $F_\gamma^{-1}(\{a\})$, and hence they are open in it. Since $F_\gamma^{-1}(\{a\})$ is homeomorphic to R, this is a contradiction.

The proof of Theorem 3.2 is done by constructing a monotone increasing sequence of domains on M^2 each of which is bounded by a geodesic biangle with certain conditions for edge angles, which is stated as follows.

Proposition 3.3. *Let γ: $[0,\infty)\to M^2$ be a ray with respect to which F_γ is non-exhaustion and F_γ does not take minimum. Let $p=\gamma(0)$ and let $t_1>0$ be taken so as to satisfy $d(\gamma(-t_1), \gamma(t_1))<2t_1$. For given monotone decreasing sequences $\{a_j\}$ and $\{\varepsilon_j\}$ so that $\lim_{j\to\infty} a_j=\inf F_\gamma$, $\lim_{j\to\infty}\varepsilon_j=0$, there exist a monotone increasing $\{s_j\}$ and geodesics σ_j: $[0,u_j]\to M^2$ and τ_j: $[0,v_j]\to M^2$ so that they fulfill (1) $\sigma_j(0)=\tau_j(0)\in F_\gamma^{-1}(\{a_j'\})$ and $\sigma_j(u_j)=\tau_j(v_j)=\gamma(s_j)$ for some $a_j'\in(a_j-\varepsilon_j, a_j+\varepsilon_j)$, (2) If D_j is the closure compact domain bounded by $\sigma_j([0,u_j])$ and $\tau_j([0,v_j])$, then $p\in D_j$ for every j and $\{D_j\}$ is a monotone increasing open cover of M^2, (3) If the angles of the geodesic biangle (σ_j,τ_j) is measured with respect to the inward pointing orientation for D_j, then $\measuredangle(-\dot\sigma_j(u_j), -\dot\tau_j(v_j))<\varepsilon_j$ and $\measuredangle(\dot\sigma_j(0), \dot\tau_j(0))<\pi$.*

Proof. For every $a_j\leqq -t_1$ we find a point z_j on $F_\gamma^{-1}\{a_j\})$ at which there are two distinct minimizing geodesics $\tilde\sigma_j,\tilde\tau_j$: $[a_j,0]\to M^2$ with $\tilde\sigma_j(a_j)=\tilde\tau_j(a_j)=z_j$ and $\tilde\sigma_j(0)$, $\tilde\tau_j(0)\in F_\gamma^{-1}(\{0\})$ and these points cut off a subarc of $F_\gamma^{-1}(\{0\})$ which contains $p\in F_\gamma^{-1}(\{0\})$ in its interior. (We see that both $\sigma_j|[a_j,\infty)$ and $\tau_j|[a_j,\infty)$ are rays asymptotic to γ.) It follows from denition of F_γ and triangle inequality that for every $\eta>0$ there is a unique ray emanating from $\tilde\sigma_j(a_j+\eta)$ and asymptotic to γ. For every j with $a_j\leqq -t_1$, we take a sequence $\{\tau_{ji}\}$ of minimizing geodesics with each geodesic τ_{ji}: $[0,v_{ji}]\to M^2$, $\|\dot q_{ji}\|=1$ and $\tau_{ji}(0)=z_j$ and $\{\tau_{ji}(v_{ji})\}$ is a divergent sequence of points on $\gamma([0,\infty))$ so that $\lim_{i\to\infty}\dot\tau_{ji}(0)=\dot{\tilde\tau}_j(0)$.

Let $r(x)$ be the convexity radius at x. For a small number $\eta \in (0, r(z_j)/4)$, we then take a sequence $\{\sigma_{ji}\}$ of minimizing geodesics σ_{ji}: $[0, u_{ji}] \to M^2$ so that $\sigma_{ji}(0) = \tilde{\sigma}_j(a_j + \eta)$, $\sigma_{ji}(u_{ji}) = \tau_{ji}(v_{ji})$. Then $\lim_{i\to\infty} \dot{\sigma}_{ji}(0) = \dot{\tilde{\sigma}}(a_j + \eta)$. Because both $\dot{\tilde{\sigma}}_j(0)$ and $\dot{\tilde{\tau}}_j(0)$ belong to the subgradient of F_γ at z_j, we see that $\measuredangle(\dot{\tilde{\sigma}}_j(0), \dot{\tilde{\tau}}_j(0)) < \pi$. And moreover for any $\varepsilon > 0$ there is an s_ε so that if $d(p, \sigma_{ji}(u_{ji})) > s_\varepsilon$, then the extension of σ_{ji} and τ_{ji} intersect at a point z'_j in a small neighborhood of z_j and the angle at the intersection differs from $\measuredangle(\dot{\tilde{\sigma}}_j(0), \dot{\tilde{\tau}}_j(0))$ by at most ε. Therefore we can choose for each j with $a_j \leqq -t_1$, an s_j in such a way that $\gamma(s_j) = \sigma_{ji}(u_{ji}) = \tau_{ji}(v_{ji})$ and $d(z_j, z'_j) < r(z_j)/4$ and: (1)′ the angle of σ_{ji} and τ_{ji} at z'_j is less than π, (2)′ $\{s_j - F_\gamma(z'_j)\}/\mathrm{Max}\{L(\sigma_{ji}), L(\tau_{ji})\} > \cos(\varepsilon_j/2)$, where $L(\sigma_{ji})$ means the length of σ_{ji} with end points z'_j and $\gamma(s_j) = \sigma_{ji}(u_{ji})$, (3)′ We denote σ_{ji} and τ_{ji} by σ_j: $[0, u_j] \to M^2$ and τ_j: $[0, v] \to M^2$ so that they are parametrized by arc length and $\sigma_j(0) = \tau_j(0) = z'_j$ and $F_\gamma(z'_j) \in (a_j - \varepsilon_j, a_j + \varepsilon_j)$. Then σ_j and τ_j cut off a subarc of $F_\gamma^{-1}(\{0\})$ whose interior contains p and (4)′ the angle $\measuredangle(\dot{\sigma}_j(0), \dot{\tau}_j(0)) < \pi$.

Let D_j be the domain bounded by $\sigma_j([0, u_j])$ and $\tau_j([0, v_j])$. From construction (1) is obviously satisfied. Since $\bar{D}_j$ is compact and its boundary intersects both $\gamma([0, \infty))$ at $\gamma(s_j)$ and $F_\gamma^{-1}(\{0\})$ at two points, the subarc of which contains p in interior, we see that $p \in D_j$. To prove that $\{D_j\}$ is monotone increasing, we first observe that the rays $\tilde{\sigma}_{j+1}$; $[a_{j+1}, \infty) \to M^2$ and $\tilde{\tau}_{j+1}$: $[a_{j+1}, \infty) \to M^2$ do not meet $\tilde{\sigma}_j([a_j, \infty))$ and $\tilde{\tau}_j([0, \infty))$, and hence the subarc of $F_\gamma^{-1}(\{0\})$ cut off by the boundary of D_{j+1} contains the subarc cut off by the boundary of D_j. Therefore s_{j+1} can be chosen so large that (1)′, (2)′, (3)′ and (4)′ are fulfilled and D_{j+1} contains D_j as a proper subset. To show $\bigcup_{j=1}^{\infty} D_j = M^2$, suppose that there is a point $y \in M^2$ which is not contained in any D_j. Let c: $[0, 1] \to M^2$ be a continuous curve joining $p = c(0)$ to $y = c(1)$. Then either $\sigma_j([0, u_j])$ or $\tau_j([0, v_j])$ intersects $c([0, 1])$ for any j. Because $\{d(\sigma_j(0), c([0, 1])\}$ and $\{d(\sigma_j(u_j), c([0, 1]))\}$ are divergent sequences either $\{\sigma_j\}$ or $\{\tau_j\}$ tends to a straight line which intersects $c([0, 1])$. But this is a contradiction.

Finally we shall prove (3). Because F_γ is differentiable on an open set of $\gamma((0, \infty))$, the gradient ∇F_γ of F_γ satisfies $\nabla F_\gamma \circ \gamma(t) = \dot{\gamma}(t)$ for all $t > 0$. Therefore we have $\langle \nabla F_\gamma(\gamma(s_j)), \dot{\sigma}_j(u_j) \rangle \geqq \{F_\gamma \circ \sigma_j(u_j) - F_\gamma \circ \sigma_j(0)\}/u_j > \cos(\varepsilon_j/2)$. This inequality means that both $\dot{\sigma}_j(u_j)$ and $\dot{\tau}_j(v_j)$ make angles with $\dot{\gamma}(s_j)$ less than $\varepsilon_j/2$. Because $\gamma([0, s_j)) \subset D_j$, we see that $-\dot{\gamma}(s_j)$ is inward pointing to D_j. Thus the edge angle of (σ_j, τ_j) at $\gamma(s_j)$ which is measured with respect to the inward pointing orientation is less than ε_j. It follows from construction of D_j that $\tilde{\sigma}_j((0, \eta)) \subset D_j$ and moreover every point on $\tilde{\sigma}_j((0, \eta))$ is joined to z'_j by a unique minimizing geodesic. This fact means that the angle $\measuredangle(\dot{\sigma}_j(0), \dot{\tau}_j(0))$ measured with respect to the inward pointing orientation is equal to the angle measured on the geodesic triangle $(z'_j, z_j, \tilde{\sigma}_j(a_j + \eta))$, which is less than π. Thus the proof is complete.

Proof of Theorem 3.2. From the previous proposition we have a family $\{D_j\}$ of monotone increasing domains so that each D_j has its boundary geodesic biangle (σ_j, τ_j). Let ε'_j and θ_j be the angles of (σ_j, τ_j) measured with respect to inward pointing orientation of D_j. Since D_j is homeomorphic to a disk, Gauss-Bonnet theorem applies to D_j to get

$$\int_{D_j} G\, d\mu + (\pi - \varepsilon'_j) + (\pi - \theta_j) = 2\pi.$$

Because $\theta_j \in (0, \pi)$ and $\varepsilon'_j \in (0, \varepsilon_j)$ we have for every j,

$$\int_{D_j} G\, d\mu < \pi + \varepsilon_j.$$

Since D_j is monotone increasing and $G \geqq 0$, the left hand side of the above inequality is monotone increasing with j. On the other hand $\pi + \varepsilon_j$ is monotone decreasing with j which tends to π. This completes the proof.

IV. Exhaustion Busemann Functions and Total Curvature

Let M^2 be as in the previous section on which G is not identically zero. Let $d\mu$ be the volume form of M^2. We shall prove here that if M^2 admits an exhaustion Busemann function then the total curvature is greater than π.

Theorem 4.1. *Assume that M^2 admits an exhaustion Busemann function. Then we have*

$$\int_{M^2} G\, d\mu > \pi.$$

Theorem 4.2. *Assume that there is a point $p \in M^2$ and a monotone increasing divergent sequence $\{t_j\}$ so that $\delta_p(t_j)/t_j < \sqrt{2}$ for all j. Then we have the following conclusions:*

(1) *There is a t_0 so that $\delta_p(t)/t \leqq \sqrt{2}$ for all $t > t_0$.*

(2) *If τ: $[0, \infty) \to M^2$ is a ray emanating from p, then F_τ is exhaustion.*

(3) *If $q \in M^2$ is any point, then there is a $t_q > 0$ so that $\delta_q(t)/t \leqq \sqrt{2}$ for all $t > t_q$.*

(4) *If σ: $[0, \infty) \to M^2$ is any ray, then F_σ is exhaustion.*

(5) $\int_{M^2} G\, d\mu > \pi.$

For the proof of the above theorems we shall prepare the following

Lemma 4.3. *Let γ: $[0, \infty) \to M^2$ be a ray with respect to which F_γ is exhaustion. Then there is a $t_0 > 0$ so that if $t > t_0$, then there exists a geodesic loop c: $[0, l] \to M^2$ with $c(0) = c(l) = \gamma(t)$.*

Proof. Let $p = \gamma(0)$. For every $t > 0$ let $S_t = \{u \in M_{\gamma(t)};\ \|u\| = t\}$ and $U \subset M_{\gamma(t)}$ be the open set such that $\bar{U} - U = C_{\gamma(t)}$. In $M_{\gamma(t)}$, the tangent cut point to $\gamma(t)$ along $\gamma((-\infty, t])$ appears beyond $-t\dot{\gamma}(t)$.

By an analogous argument developed in Lemma 1.3, we find a $t_0 > 0$ so that if $t > t_0$ then $-t\dot{\gamma}(t)$ and $t\dot{\gamma}(t)$ do not belong to the same component of $S_t \cap \bar{U}$. In fact suppose that there is a divergent sequence $\{t_j\}$ in such a way that $-t_j\dot{\gamma}(t_j)$ and $t_j\dot{\gamma}(t_j)$ belong to the same component of $S_{t_j} \cap \bar{U}$. Let $R > 0$ be any fixed number. Then for every j with $t_j > R$, the exponentiated image of the half t_j-circle sketched in $S_{t_j} \cap \bar{U}$ centered at the origin and with the endpoints $-t_j\dot{\gamma}(t_j)$ and $t_j\dot{\gamma}(t_j)$ is a continuous curve in M^2. Therefore we find a u_j on this half circle so that if $x_j = \exp_{\gamma(t_j)} u_j$, then $d(p, x_j) = R$. Because $\{x_j\}$ is a bounded sequence, there is a point x on M^2 to which a subsequence of it converges. From definition of F_γ it follows that $x\in F_\gamma^{-1}(\{0\})$ and $d(p,x)=R$. Since R is any, this means that $F_\gamma^{-1}(\{0\})$ is not compact, contradicting that F_γ is exhaustion.

Let $t>t_0$ be fixed and let u, $v\in S_t\cap \bar{U}$ be the boundary of the component of $S_t\cap \bar{U}$ which contains $-t\dot\gamma(t)$ in its interior. Let $K(t)$ be the component of $M^2-B_t(\gamma(t))$ which contains the cut point of $\gamma(t)$ along $\gamma((-\infty,t])$. $K(t)$ is compact because there is a unique unbounded component of $M^2-B_t(\gamma(t))$, which contains $\gamma([2t,\infty))$. The boundary of $K(t)$ is the exponentiated image of the component of $S_t\cap \bar{U}$ which contains $-t\dot\gamma(t)$ in its interior. We see that the boundary of $K(t)$ is homeomorphic to S^1 and converges uniformly to $F_\gamma^{-1}(\{0\})$ as $t\to\infty$. And hence we have $\exp_{\gamma(t)}u=\exp_{\gamma(t)}v\in C(\gamma(t))$, and $-t\dot\gamma(t)$ is obtained by a convex combination of u and v with a positive coefficient. Indeed let c_1,c_2: $[0,t]\to M^2$ be the minimizing geodesics with $c_1(s)=\exp_{\gamma(t)}\mathrm{s}\mathrm{u}/\|\mathrm{u}\|$, and $c_2(s)=\exp_{\gamma(t)}sv/\|v\|$. From $t-d(x,\gamma(t))\leqq F_\gamma(x)$ for any $x\in M^2$, we see that $F_\gamma(c_i(s))\in[0,t]$ holds for all $i=1,2$ and all $s\in[0,t]$. Since F_γ is differentiable at $\gamma(t)$ and $\nabla F_\gamma(\gamma(t))=\dot\gamma(t)$, and since $F_\gamma\circ c_i(s)$ is strictly monotone decreasing, we have $\langle \nabla F_\gamma, \dot c_i(0)\rangle<0$ for $i=1,2$.

Let c: $[0,2t]\to M^2$ be the broken geodesic obtained by $c(s)=c_1(s)$ for $s\in[0,t]$ and $c(s)=c_2(2t-s)$ for $s\in[t,2t]$. Let D be the domain bounded by $c([0,2t])$. Then D is homeomorphic to a 2-disl. Since $-\dot\gamma(t)$ is inward pointing to D, we have $p\in D$. Because $F_\gamma^{-1}(\{0\})$ intersects $c([0,2t])$ (if the intersection exists) only at $c(t)$ and $p\in D$, we have $F_\gamma^{-1}(\{0\})\in\bar{D}$, and hence $K\subset\bar{D}$. Because $F_\gamma\circ c_i$ is monotone decreasing, we see that $\dot c_i(t)$ is inward pointing to D. Hence the edge angle at this point is greater than π if it is measured with respect to the inward pointing orientation.

We shall assert that c is homotopic to the circle $F_\gamma^{-1}(\{0\})$. By means of the basic construction and Proposition 1.3 of [3], we see that

$$\begin{aligned}&F_\gamma^{-1}(\{t-kr\})\\&\quad=\{x\in F_\gamma^{-1}((-\infty,t-(k-1)r]):\ d(x,F_\gamma^{-1}(\{t-(k-1)r\})=r\}\qquad\text{for } k\\&\quad=0,1,\dots,n,\end{aligned}$$

where n and r are taken so that r is less than half of the convexity radius of the compact set $F_\gamma^{-1}((-\infty,t])$ and $t=nr$. For each k, $F_\gamma^{-1}([t-kr,t-(k-1)r])$ is homeomorphic to $F_\gamma^{-1}(\{t-(k-1)r\})\times[0,1]$ and the homeomorphism is obtained by "perpendicular" geodesics from points on $F_\gamma^{-1}(t-kr,t-(k-1)r])$ to $F_\gamma^{-1}((-\infty,t-(k+1)r])$, where a "perpendicular" geodesic from x to a closed convex set C is by definition a minimizing geodesic from x to a point on C whose length realizes the distance $d(x,C)$. Because $F_\gamma\circ c_1$ and $F_\gamma\circ c_2$ are strictly monotone decreasing, each of c_i has a unique intersection with levels of F_γ. Moreover every point of $\gamma([0,t])$ is mapped by this homeomorphism into itself. Hence there is a subinterval I_1,I_2 of 0 in $[0,t]$ so that $c_i(I_i)$ is homeomorphically deformed into $F_\gamma^{-1}(\{t-r\})$ via the perpendicular geodesics to $F_\gamma^{-1}((-\infty,t-2r])$ and by which $\gamma(t)$ is mapped to $\gamma(t-r)$, and ends of which are fixed because F_γ takes value $t-r$ at the ends. By iterating this procedure n times we obtain a homotopy between c and $F_\gamma^{-1}(\{0\})$. We note that the image of this homotopy is in $F_\gamma^{-1}([0,t])$.

We shall finally find a geodisc loop at $\gamma(t)$ whose image is in $F_\gamma^{-1}([0,t])$ and which is homotopic to c. Let Y be a piecewise smooth vector field along c which is smooth on $[0,t]$ and on $[t,2t]$ and $Y(0)=0$ and $Y(s)$ is pointing out with respect to D. Then the 1-parameter variation associated with Y has negative first variation because the edge angle of c at $c(t)$ is greater than π. Thus we find a geodesic loop at $\gamma(t)$ which has minimal length among loops homotopic to c in the set $F_\gamma^{-1}([0,t])$. Thus the proof of Lemma 4.3 is complete.

Proof of Theorem 4.1. Let $p=\gamma(0)$ and let t_0 be obtained in Lemma 4.3 for γ and fix a large $t>t_0$. There is a geodesic loop $\sigma\colon [0,l]\to F_\gamma^{-1}([0,t])$ so that $\sigma(0)=\sigma(l)$ $=\gamma(t)$. Since M^2 is diffeomorphic to R^2, σ bounds a domain D^* which is homeomorphic to a 2-disk. Then D^* contains D which is obtained in the proof of Lemma 4.3 and which contains p. Therefore $-\dot\gamma(t)$ is inward pointing with respect to D^*. Because $F_\gamma\circ\sigma\colon [0,l]\to R$ is convex and $\nabla F_\gamma\circ\gamma(t)=\dot\gamma(t)$, we observe that $\langle\dot\gamma(t),\dot\sigma(0)\rangle\leqq 0$ and $\langle\dot\gamma(t),-\dot\sigma(l)\rangle\leqq 0$. This fact implies that the angle at $\sigma(0)$ measured with respect to the inward pointing orientation is equal to $\sphericalangle(\dot\sigma(0),$ $-\dot\gamma(t))+\sphericalangle(-\dot\sigma(l),-\dot\gamma(t))\leqq\pi$. Thus we have $\sphericalangle(\dot\sigma(0),\dot\sigma(l))\in[0,\pi)$ with respect to the inward pointing orientation. We only apply Gauss-Bonnet theorem to obtain

$$\int_{D^*} G\,d\mu+\sphericalangle(\dot\sigma(0),\dot\sigma(l))=2\pi.$$

Therefore we obtain

$$\int_{M^2} G\,d\mu\geqq\int_{D^*} G\,d\mu=2\pi-\sphericalangle(\dot\sigma(0),\dot\sigma(l))>\pi.$$

This completes the proof of Theorem 4.1.

Proof of Theorem 4.2. Suppose that there is a divergent sequence $\{u_j\}$ so that $\delta_p(u_j)/u_j>\sqrt{2}$. Then Theorem 2.4 implies that $\delta_p(t)/t\geqq\sqrt{2}$ for all $t>0$, a contradiction. This proves (1). Suppose that there is a ray $\tau\colon [0,\infty)\to M^2$ emanating from $p=\tau(0)$ so that F_τ is not exhaustion. Then Proposition 2.2 and (1) in Theorem 3.1 imply that $\delta_p(t)/t>\sqrt{2}$ for all $t>0$, a contradiction. This proves (2). To prove (3) and (4) let $q\in M^2$ be any point and apply Theorem 1.2 to get a monotone increasing divergent sequence $\{u_j\}$ so that

$$\sqrt{2}\geqq\lim_{j\to\infty}\delta_p(u_j)/u_j\geqq\limsup_{j\to\infty}\delta_q(u_j)/u_j.$$

Thus there is a number $t_q>0$ so that $\delta_q(u_j)/u_j\leqq\sqrt{2}$ for all j with $u_j>t_q$. Suppose that there is a ray $\sigma\colon [0,\infty)\to M^2$ emanating from $q=\sigma(0)$ so that F_σ is non-exhaustion. Then Proposition 2.2 and (1) in Theorem 3.1 imply that $\delta_q(t)/t>\sqrt{2}$ for all $t>0$, a contradiction. If $\delta_q(v_j)/v_j>\sqrt{2}$ for some monotone increasing divergent sequence, then there is a ray $\sigma\colon [0,\infty)\to M^2$ so that F_σ is non-exhaustion, a contradiction. Now (5) is obvious from (2) and Theorem 4.1. This completes the proof of Theorem 4.2.

Summing up the results obtained in §§3 and 4, we have the following

Theorem 4.4. *Let M^2 be a connected, oriented, complete and noncompact Riemannian 2-manifold without boundary. Let the Gaussian curvature of M^2 be nonnegative everywhere. Then the total curvature of M^2 is greater than π if and only if every Busemann function is exhaustion, and the total curvature of M^2 is not greater than π if and only if every Busemann function is non-exhaustion.*

As a direct consequence of the above theorem, we have the following

Corollary to Theorem 4.4. *Let M^2 be a connected, complete and noncompact Riemannian 2-manifold without boundary and let the Gaussian curvature of M^2 is nonnegative everywhere. If there is an exhaustion Busemann function on M^2, then all Busemann functions on it are exhaustion, and if there is a non-exhaustion Busemann function on M^2, then all Busemann functions are non-exhaustion.*

Proof. If M^2 is oriented, then the conclusion is obvious from Theorem 4.4. Let M^2 be non-orientable. Then a classification theorem of complete noncompact Riemannian 2-manifolds which admit locally nonconstant convex functions implies that M^2 is in this case isometric to a flat open Mobius strip. Therefore in this case all Busemann functions on it are exhaustion. This completes the proof of Corollary.

The author would like to express his thanks to the referee who pointed several errors in the original manuscript out to him.

References

1. Bangert, V.: Riemannsche Mannigfaltigkeiten mit nicht konstanter konvexer Funktionen, Arch. Math., **31**, 163–170 (1978)
2. Busemann, H.: The geometry of geodesics, New York: Academic Press INC. 1955
3. Cheeger, J., Gromoll, D.: On the structure of complete manifolds of nonnegative curvature, Ann. of Math., **96**, 413–443 (1972)
4. Cheeger, J., Ebin, D.: Comparison theorems in riemannan geometry, Amsterdam-Oxford-New York: North Holland Math. Library 9, 1975
5. Cohn-Vossen, S.: Kürzeste Wege und Totalkrümmung auf Flächen, Comp. Math., **2**, 69–133 (1935)
6. Cohn-Vossen, S.: Totalkrümmung und geodätische Linien auf einfachzusammehängenden offenen vollständigen Flächenstücken, Recueil Math. Moskau, **43**, 139–163 (1936)
7. Greene, R., Shiohama, K.: Convex functions on complete noncompact manifolds, Preprint at UCLA, 1978
8. Greene, R., Wu, H.: C^∞ convex functions and manifolds of positive curvature, Acta Math., **137**, 209–245 (1976)
9. Gromoll, D., Meyer, W.: On complete open manifolds of positive curvature, Ann. of Math., **90**, 75–90 (1969)
10. Maeda, M.: On the total curvature of noncompact Riemannian manifolds, Kodai Math. Sem. Rep., **26**, 95–99 (1974)
11. Poor, W.: Some results on nonnegatively curved manifolds, J. Differential Geometry, **9**, 583–600 (1974)
12. Shiohama, K.: An extension of a theorem of Myers, J. Math. Soc. Japan, **27**, 561–569 (1975)
13. Toponogov, A.: Riemannian spaces having their curvature bounded below by a positive number, Amer. Math. Soc. Transl., **37**, 291–336 (1964)
14. Toponogov, A., Riemannian spaces which contain straight lines, Amer. Math. Soc. Transl. **37**, 287–290 (1964)

Received March 28/Revised June 11, 1979