Documenta Mathematica

Journal der Deutschen Mathematiker-Vereinigung

Band 2

1997

ISSN 1431-0635 Print ISSN 1431-0643 Internet

Documenta Mathematica, Journal der Deutschen Mathematiker-Vereinigung, veröffentlicht Forschungsarbeiten aus allen mathematischen Gebieten und wird in traditioneller Weise referiert.

Documenta Mathematica erscheint am World Wide Web unter der Adresse:

`http://www.mathematik.uni-bielefeld.de/documenta`

Artikel können als TEX-Dateien per E-Mail bei einem der Herausgeber eingereicht werden. Hinweise für die Vorbereitung der Artikel können unter der obigen WWW-Adresse gefunden werden.

Documenta Mathematica, Journal der Deutschen Mathematiker-Vereinigung, publishes research manuscripts out of all mathematical fields and is refereed in the traditional manner.

Documenta Mathematica is published on the World Wide Web under the address:

`http://www.mathematik.uni-bielefeld.de/documenta`

Manuscripts should be submitted as TEX files by e-mail to one of the editors. Hints for manuscript preparation can be found under the above WWW-address.

ISSN 1431-0635 Documenta Mathematica (Print)
ISSN 1431-0643 Documenta Mathematica (Internet)

Anschrift des technischen geschäftsführenden Herausgebers:
Ulf Rehmann, Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld

Documenta Mathematica

Journal der Deutschen Mathematiker-Vereinigung

Band 2, 1997

On the Approximation Numbers of Large Toeplitz Matrices

A. Böttcher*

Received: January 14, 1997

Communicated by Alfred K. Louis

Abstract. The kth approximation number $s_k^{(p)}(A_n)$ of a complex $n \times n$ matrix A_n is defined as the distance of A_n to the $n \times n$ matrices of rank at most $n-k$. The distance is measured in the matrix norm associated with the l^p norm $(1 < p < \infty)$ on $\mathbf{C}^n$. In the case $p=2$, the approximation numbers coincide with the singular values.

We establish several properties of $s_k^{(p)}(A_n)$ provided A_n is the $n \times n$ truncation of an infinite Toeplitz matrix A and n is large. As $n \to \infty$, the behavior of $s_k^{(p)}(A_n)$ depends heavily on the Fredholm properties (and, in particular, on the index) of A on l^p.

This paper is also an introduction to the topic. It contains a concise history of the problem and alternative proofs of the theorem by G. Heinig and F. Hellinger as well as of the scalar-valued version of some recent results by S. Roch and B. Silbermann concerning block Toeplitz matrices on l^2.

1991 Mathematics Subject Classification: Primary 47B35; Secondary 15A09, 15A18, 15A60, 47A75, 47A58, 47N50, 65F35

1. Introduction

Throughout this paper we tacitly identify a complex $n \times n$ matrix with the operator it induces on $\mathbf{C}^n$. For $1 < p < \infty$, we denote by $\mathbf{C}_p^n$ the space $\mathbf{C}^n$ with the l^p norm,

$$\|x\|_p := \Big(|x_1|^p + \ldots + |x_n|^p\Big)^{1/p},$$

and given a complex $n \times n$ matrix A_n, we put

$$\|A_n\|_p := \sup_{x \neq 0} \Big(\|A_n x\|_p / \|x\|_p\Big). \tag{1}$$

*Research supported by the Alfried Krupp Förderpreis für junge Hochschullehrer of the Krupp Foundation

We let $\mathcal{B}(\mathbf{C}_p^n)$ stand for the Banach algebra of all complex $n \times n$ matrices with the norm (1). For $j \in \{0, 1, \ldots, n\}$, let $\mathcal{F}_j^{(n)}$ be the collection of all complex $n \times n$ matrices of rank at most j, i.e., let

$$\mathcal{F}_j^{(n)} := \Big\{ F \in \mathcal{B}(\mathbf{C}_p^n) : \dim \operatorname{Im} F \le j \Big\}.$$

The kth approximation number ($k \in \{0, 1, \ldots, n\}$) of $A_n \in \mathcal{B}(\mathbf{C}_p^n)$ is defined as

$$s_k^{(p)}(A_n) := \operatorname{dist}(A_n, \mathcal{F}_{n-k}^{(n)}) := \min \Big\{ \|A_n - F_n\|_p : F_n \in \mathcal{F}_{n-k}^{(n)} \Big\}. \tag{2}$$

(note that $\mathcal{F}_j^{(n)}$ is a closed subset of $\mathcal{B}(\mathbf{C}_p^n)$). Clearly,

$$0 = s_0^{(p)}(A_n) \le s_1^{(p)}(A_n) \le \ldots \le s_n^{(p)}(A_n) = \|A_n\|_p.$$

It is easy to show (see Proposition 9.2) that

$$s_1^{(p)}(A_n) = \begin{cases} 1/\|A_n^{-1}\|_p & \text{if} \quad A_n \quad \text{is invertible,} \\ 0 & \text{if} \quad A_n \quad \text{is not invertible.} \end{cases} \tag{3}$$

Notice also that in the case $p = 2$ the approximation numbers $s_1^{(2)}(A_n), \ldots, s_n^{(2)}(A_n)$ are just the singular values of A_n, i.e., the eigenvalues of $(A_n^* A_n)^{1/2}$.

Let $\mathbf{T}$ be the complex unit circle and let $a \in L^\infty := L^\infty(\mathbf{T})$. The $n \times n$ Toeplitz matrix $T_n(a)$ generated by a is the matrix

$$T_n(a) := (a_{j-k})_{j,k=1}^n \tag{4}$$

where a_l ($l \in \mathbf{Z}$) is the lth Fourier coefficient of a,

$$a_l := \frac{1}{2\pi} \int_0^{2\pi} a(e^{i\theta}) e^{-il\theta} \, d\theta.$$

This paper is devoted to the limiting behavior of the numbers $s_k^{(p)}(T_n(a))$ as n goes to infinity.

Of course, the study of properties of $T_n(a)$ as $n \to \infty$ leads to the consideration of the infinite Toeplitz matrix

$$T(a) := (a_{j-k})_{j,k=1}^\infty.$$

The latter matrix induces a bounded operator on $l^2 := l^2(\mathbf{N})$ if (and only if) $a \in L^\infty$. Acting with $T(a)$ on $l^p := l^p(\mathbf{N})$ is connected with a multiplier problem in case $p \ne 2$. We let M_p stand for the set of all $a \in L^\infty$ for which $T(a)$ generates a bounded operator on l^p. The norm of this operator is denoted by $\|T(a)\|_p$. The function a is usually referred to as the symbol of $T(a)$ and $T_n(a)$.

In this paper, we prove the following results.

Theorem 1.1. *If $a \in M_p$ then for each k,*

$$s_{n-k}^{(p)}\Big(T_n(a)\Big) \to \|T(a)\|_p \quad \text{as} \quad n \to \infty.$$

Theorem 1.2. *If $a \in M_p$ and $T(a)$ is not normally solvable on l^p then for each k,*

$$s_k^{(p)}\Big(T_n(a)\Big) \to 0 \text{ as } n \to \infty$$

Let $M_{\langle 2\rangle} := L^\infty$. For $p \neq 2$, we define $M_{\langle p\rangle}$ as the set of all functions $a \in L^\infty$ which belong to $M_{\tilde{p}}$ for all $\tilde{p}$ in some open neighborhood of p (which may depend on a). A well known result by Stechkin says that $a \in M_p$ for all $p \in (1,\infty)$ whenever $a \in L^\infty$ and the total variation $V_1(a)$ of a is finite and that in this case

$$\|T(a)\|_p \leq C_p\Big(\|a\|_\infty + V_1(a)\Big) \tag{5}$$

with some constant $C_p < \infty$ (see, e.g., [5, Section 2.5(f)] for a proof). We denote by PC the closed subalgebra of L^∞ constituted by all piecewise continuous functions. Thus, $a \in PC$ if and only if $a \in L^\infty$ and the one-sided limits

$$a(t \pm 0) := \lim_{\varepsilon \to 0\pm 0} a(e^{i(\theta+\varepsilon)})$$

exist for every $t = e^{i\theta} \in \mathbf{T}$. By virtue of (5), the intersection $PC \cap M_{\langle p\rangle}$ contains all piecewise continuous functions of finite total variation.

Throughout what follows we define $q \in (1,\infty)$ by $1/p + 1/q = 1$ and we put

$$[p,q] := \Big[\min\{p,q\}, \max\{p,q\}\Big].$$

One can show that if $a \in M_p$, then $a \in M_r$ for all $r \in [p,q]$ (see, e.g., [5, Section 2.5(c)]).

Here is the main result of this paper.

Theorem 1.3. *Let a be a function in $PC \cap M_{\langle p\rangle}$ and suppose $T(a)$ is Fredholm of the same index $-k$ ($\in \mathbf{Z}$) on l^r for all $r \in [p,q]$. Then*

$$\lim_{n\to\infty} s_{|k|}^{(p)}\Big(T_n(a)\Big) = 0 \quad \text{and} \quad \liminf_{n\to\infty} s_{|k|+1}^{(p)}\Big(T_n(a)\Big) > 0.$$

For $p = 2$, Theorems 1.2 and 1.3 are special cases of results by Roch and Silbermann [20], [21]. Since a Toeplitz operator on l^2 with a piecewise continuous symbol is either Fredholm (of some index) or not normally solvable, Theorems 1.2 and 1.3 completely identify the approximation numbers (= singular values) which go to zero in the case $p = 2$.

Now suppose $p \neq 2$. If $a \in C \cap M_{\langle p\rangle}$, then $T(a)$ is again either Fredholm or not normally solvable, and hence Theorems 1.2 and 1.3 are all we need to see which approximation numbers converge to zero. In the case where $a \in PC \cap M_{\langle p\rangle}$ we have three mutually excluding possibilities (see Section 3):

(i) $T(a)$ is Fredholm of the same index $-k$ on l^r for all $r \in [p,q]$;

(ii) $T(a)$ is not normally solvable on l^p or not normally solvable on l^q;

(iii) $T(a)$ is normally solvable on l^p and l^q but not normally solvable on l^r for some $r \in (p,q) := [p,q] \setminus \{p,q\}$.

In the case (i) we can apply Theorem 1.3. Since

$$s_k^{(p)}\Big(T_n(a)\Big) = s_k^{(q)}\Big(T_n(a)\Big) \tag{6}$$

(see (35)), Theorem 1.2 disposes of the case (ii). I have not been able to settle the case (iii). My conjecture is as follows.

Conjecture 1.4. *In the case* (iii) *we have*

$$s_k^{(p)}\Big(T_n(a)\Big) \to 0 \ \text{ as } \ n \to \infty$$

for every fixed k.

The paper is organized as follows. Section 2 is an attempt at presenting a short history of the topic. In Section 3 we assemble some results on Toeplitz operators on l^p which are needed to prove the three theorems stated above. Their proofs are given in Sections 4 to 6. The intention of Sections 7 and 8 is to illustrate how some simple constructions show a very easy way to understand the *nature* of the Heinig/Hellinger and Roch/Silbermann results. Notice, however, that the approach of Sections 7 and 8 cannot replace the methods of these authors. They developed some sort of high technology which enabled them to tackle the block case and more general approximation methods, while in these two sections it is merely demonstrated that in the scalar case (almost) all problems can be solved with the help of a few crowbars (Theorems 7.1, 7.2, 7.4). Nevertheless, beginners will perhaps appreciate reading Sections 7 and 8 before turning to the papers [13] and [25], [20].

2. Brief history

The history of the lowest approximation number $s_1^{(p)}(T_n(a))$ is the history of the finite section method for Toeplitz operators: by virtue of (3), we have

$$s_1^{(p)}\Big(T_n(a)\Big) \to 0 \Longleftrightarrow \|T_n^{-1}(a)\|_p \to \infty.$$

We denote by $\Phi_k(l^p)$ the collection of all Fredholm operators of index k on l^p. The equivalence

$$\limsup_{n\to\infty} \|T_n^{-1}(a)\|_p < \infty \Longleftrightarrow T(a) \in \Phi_0(l^p) \tag{7}$$

was proved by Gohberg and Feldman [7] in two cases: if $a \in C \cap M_{\langle p \rangle}$ (where C stands for the continuous functions on $\mathbf{T}$) or if $p = 2$ and $a \in PC$. For $a \in PC \cap M_{\langle p \rangle}$, the equivalence

$$\limsup_{n\to\infty} \|T_n^{-1}(a)\|_p < \infty \Longleftrightarrow T(a) \in \Phi_0(l^r) \ \text{ for all } \ r \in [p,q] \tag{8}$$

holds. This was shown by Verbitsky and Krupnik [30] in the case where a has a single jump, by Silbermann and the author [3] for symbols with finitely many jumps, and finally by Silbermann [23] for symbols with a countable number of jumps. In the work of many authors, including Ambartsumyan, Devinatz, Shinbrot, Widom, Silbermann, it was pointed out that (7) is also true if

$$p = 2 \text{ and } a \in (C + H^\infty) \cup (C + \overline{H^\infty}) \cup PQC$$

(see [4], [5]). Also notice that the implication "$\Longrightarrow$" of (8) is valid for every $a \in M_p$. Treil [26] proved that there exist symbols $a \in M_{\langle 2 \rangle} = L^\infty$ such that $T(a) \in \Phi_0(l^2)$ but $\|T_n^{-1}(a)\|_2$ is not uniformly bounded; concrete symbols with this property can be found in the recent article [2, Section 7.7].

The Toeplitz matrices

$$T_n(\varphi_\gamma) = \left(\frac{1}{j - k + \gamma} \right)_{j,k=1}^n \quad (\gamma \notin \mathbf{Z})$$

are the elementary building blocks of general Toeplitz matrices with piecewise continuous symbols and have therefore been studied for some decades. The symbol is given by

$$\varphi_\gamma(e^{i\theta}) = \frac{\pi}{\sin \pi\gamma} e^{i\pi\gamma} e^{-i\gamma\theta}, \quad \theta \in [0, 2\pi).$$

This is a function in PC with a single jump at $e^{i\theta} = 1$. Tyrtyshnikov [27] focussed attention on the singular values of $T_n(\varphi_\gamma)$. He showed that

$$s_1^{(2)}\Big(T_n(\varphi_\gamma)\Big) = O(1/n^{|\gamma| - 1/2}) \text{ if } \gamma \in \mathbf{R} \text{ and } |\gamma| > 1/2$$

and that there are constants $c_1, c_2 \in (0, \infty)$ such that

$$c_1 / \log n \le s_1^{(2)}\Big(T_n(\varphi_{1/2})\Big) \le c_2 / \log n.$$

Curiously, the case $|\gamma| < 1/2$ was left as an open problem in [27], although from the standard theory of Toeplitz operators with piecewise continuous symbols it is well known that

$$T(\varphi_\gamma) \in \Phi_0(l^2) \Longleftrightarrow |\mathrm{Re}\,\gamma| < 1/2$$

(see, e.g., [7, Theorem IV.2.1] or [5, Proposition 6.24]), which together with (7) (for $p = 2$ and $a \in PC$) implies that

$$\liminf_{n \to \infty} s_1^{(2)}\Big(T_n(\varphi_\gamma)\Big) = 0 \text{ if } |\mathrm{Re}\,\gamma| \ge 1/2 \tag{9}$$

and

$$\liminf_{n \to \infty} s_1^{(2)}\Big(T_n(\varphi_\gamma)\Big) > 0 \text{ if } |\mathrm{Re}\,\gamma| < 1/2$$

(see [20]). A simple and well known argument (see the end of Section 3) shows that in (9) the liminf can actually be replaced by lim.

Also notice that it was already in the seventies when Verbitsky and Krupnik [30] proved that

$$\lim_{n \to \infty} s_1^{(p)}\Big(T_n(\varphi_\gamma)\Big) = 0 \iff |\mathrm{Re}\,\gamma| \ge \min\{1/p, 1/q\}$$

(full proofs are also in [4, Proposition 3.11] and [5, Theorem 7.37; in part (iii) of that theorem there is a misprint: the $-1/p < \operatorname{Re}\beta < 1/q$ must be replaced by $-1/q < \operatorname{Re}\beta < 1/p$]).

As far as I know, collective phenomena of $s_1^{(p)}(T_n(a)), \dots, s_n^{(p)}(T_n(a))$ have been studied only for $p = 2$, and throughout the rest of this section we abbreviate $s_k^{(2)}(T_n(a))$ to $s_k(T_n(a))$.

In 1920, Szegö showed that if $a \in L^\infty$ is real-valued and F is continuous on $\mathbf{R}$, then

$$\frac{1}{n}\sum_{k=1}^{n} F\Big(s_k(T_n(a))\Big) \to \frac{1}{2\pi}\int_0^{2\pi} F\Big(|a(e^{i\theta})|\Big)\, d\theta. \tag{10}$$

In the eighties, Parter [15] and Avram [1] extended this result to arbitrary (complex-valued) symbols $a \in L^\infty$. Formula (10) implies that

$$\Big\{s_k(T_n(a))\Big\}_{k=1}^{n} \text{ and } \Big\{|a(e^{2\pi ik/n})|\Big\}_{k=1}^{n} \tag{11}$$

are equally distributed (see [9] and [29]).

Research into the asymptotic distribution of the singular values of Toeplitz matrices was strongly motivated by a phenomenon discovered by C. Moler in the middle of the eighties. Moler observed that almost all singular values of $T_n(\varphi_{1/2})$ are concentrated in $[\pi - \varepsilon, \pi]$ where ε is very small. Formula (10) provides a way to understand this phenomenon: letting $F = 1$ on $[0, \pi - 2\varepsilon]$ and $F = 0$ on $[\pi - \varepsilon, \pi]$ and taking into account that $|\varphi_{1/2}| = 1$, one gets

$$\frac{1}{n}\sum_{k=1}^{n} F\Big(s_k(T_n(\varphi_{1/2}))\Big) \to \frac{1}{2\pi}\int_0^{2\pi} F(1)\, d\theta = F(1) = 0,$$

which shows that the percentage of the singular values of $T_n(\varphi_{1/2})$ which are located in $[0, \pi - 2\varepsilon]$ goes to zero as n increases to infinity.

Widom [32] was the first to establish a second order result on the asymptotics of singular values. Under the assumption that

$$a \in L^\infty \text{ and } \sum_{n\in\mathbf{Z}} |n|\,|a_n|^2 < \infty$$

and that $F \in C^3(\mathbf{R})$, he showed that

$$\sum_{k=1}^{n} F\Big(s_k^2(T_n(a))\Big) = \frac{n}{2\pi}\int_0^{2\pi} F\Big(|a(e^{i\theta})|^2\Big)\, d\theta + E_F(a) + o(1)$$

with some constant $E_F(a)$, and he gave an expression for $E_F(a)$. He also introduced two limiting sets of the sets

$$\Sigma(T_n(a)) := \Big\{s_1(T_n(a)), \dots, s_n(T_n(a))\Big\},$$

which, following the terminology of [19], are defined by

$$\begin{aligned}\Lambda_{\text{part}}\left(\Sigma(T_n(a))\right) &:= \{\lambda\in\mathbf{R}:\lambda \text{ is partial limit of some sequence}\\ &\qquad\qquad \{\lambda_n\} \text{ with } \lambda_n\in\Sigma(T_n(a))\},\\ \Lambda_{\text{unif}}\left(\Sigma(T_n(a))\right) &:= \{\lambda\in\mathbf{R}:\lambda \text{ is the limit of some sequence}\\ &\qquad\qquad \{\lambda_n\} \text{ with } \lambda_n\in\Sigma(T_n(a))\}.\end{aligned}$$

It turned out that for large classes of symbols a we have

$$\Lambda_{\text{part}}\left(\Sigma(T_n(a))\right)=\Lambda_{\text{unif}}\left(\Sigma(T_n(a))\right)=\operatorname{sp}\left(T(\overline{a})T(a)\right)^{1/2} \tag{12}$$

where $\operatorname{sp} A := \{\lambda\in\mathbf{C}: A-\lambda I \text{ is not invertible}\}$ denotes the spectrum of A (on l^2) and $\overline{a}$ is defined by $\overline{a}(e^{i\theta}) := \overline{a(e^{i\theta})}$. Note that $T(\overline{a})$ is nothing but the adjoint $T^*(a)$ of $T(a)$. Widom [32] proved (12) under the hypothesis that $a\in PC$ or that a is locally self-adjoint, while Silbermann [24] derived (12) for locally normal symbols. Notice that symbols in PC or even in PQC are locally normal.

In the nineties, Tyrtyshnikov [28], [29] succeeded in proving that the sets (11) are equally distributed under the sole assumption that $a\in L^2 := L^2(\mathbf{T})$. His approach is based on the observation that if $\|A_n-B_n\|_F = o(n)$, where $\|\cdot\|_F$ stands for the Frobenius (or Hilbert-Schmidt) norm, then A_n and B_n have equally distributed singular values. The result mentioned can be shown by taking $A_n = T_n(a)$ and choosing appropriate circulants for B_n.

The development received a new impetus from Heinig and Hellinger's 1994 paper [13]. They considered normally solvable Toeplitz operators on l^2 and studied the problem whether the Moore-Penrose inverses of $T_n^+(a)$ of $T_n(a)$ converge strongly on l^2 to the Moore-Penrose inverse $T^+(a)$ of $T(a)$. Recall that the Moore-Penrose inverse of a normally solvable Hilbert space operator A is the (uniquely determined) operator A^+ satisfying

$$AA^+A=A,\;\; A^+AA^+=A^+,\;\; (A^+A)^*=A^+A,\;\; (AA^+)^*=AA^+.$$

If $a\in C$, then $T(a)$ is normally solvable on l^2 if and only if $a(t)\neq 0$ for all $t\in\mathbf{T}$. When writing $T_n^+(a)\to T^+(a)$, we actually mean that $T_n^+(a)P_n\to T^+(a)$, where P_n is the projection defined by

$$P_n:\{x_1,x_2,x_3,\ldots\}\mapsto\{x_1,x_2,\ldots,x_n,0,0,\ldots\}. \tag{13}$$

It is not difficult to verify that $T_n^+(a)\to T^+(a)$ strongly on l^2 if and only if $T(a)$ is normally solvable and

$$\limsup_{n\to\infty}\|T_n^+(a)\|_2<\infty. \tag{14}$$

Heinig and Hellinger investigated normally solvable Toeplitz operators $T(a)$ with symbols in the Wiener algebra W,

$$a\in W \Longleftrightarrow \|a\|_W := \sum_{n\in\mathbf{Z}}|a_n|<\infty,$$

and they showed that then (14) is satisfied if and only if there is an $n_0\geq 1$ such that

$$\operatorname{Ker} T(a)\subset\operatorname{Im}P_{n_0} \;\text{ and }\; \operatorname{Ker}T(\overline{a})\subset\operatorname{Im}P_{n_0}, \tag{15}$$

where $\operatorname{Ker} A := \{x \in l^2 : Ax = 0\}$ and $\operatorname{Im} A := \{Ax : x \in l^2\}$. (This formulation of the Heinig-Hellinger result is due to Silbermann [25].) Conditions (15) are obviously met if $T(a)$ is invertible, in which case even $\|T_n^{-1}(a)\|_2$ is uniformly bounded. The really interesting case is the one in which $T(a)$ is not invertible, and in that case (15) and thus (14) are highly instable. For example, if a is a rational function (without poles on $\mathbf{T}$) and $\lambda \in \operatorname{sp} T(a)$, then

$$\limsup_{n\to\infty} \|T_n^+(a-\lambda)\|_2 < \infty \tag{16}$$

can only hold if λ belongs to $\operatorname{sp} T_n(a)$ for all sufficiently large n. Consequently, (16) implies that λ lies in $\Lambda_{\mathrm{unif}}(\operatorname{sp} T_n(a))$, and the latter set is extremely "thin": it is contained in a finite union of analytic arcs (see [22] and [6]).

What has Moore-Penrose invertibility to do with singular values ? The answer is as follows: if $A_n \in \mathcal{B}(\mathbf{C}_2^n)$ and $s_k(A_n)$ is the smallest nonzero singular value of A_n, then

$$\|A_n^+\|_2 = 1/s_k(A_n).$$

Thus, (14) holds exactly if there exists a $d > 0$ such that

$$\Sigma(T_n(a)) \subset \{0\} \cup [d, \infty) \tag{17}$$

for all sufficiently large n.

Now Silbermann enters the scene. He replaced the Heinig-Hellinger problem by another one. Namely, given $T(a)$, is there a sequence $\{B_n\}$ of operators $B_n \in \mathcal{B}(\mathbf{C}_2^n)$ with the following properties: there exists a bounded operator B on l^2 such that

$$B_n \to B \;\text{ and }\; B_n^* \to B^* \;\text{ strongly on }\; l^2$$

and

$$\begin{aligned}&\|T_n(a)B_nT_n(a) - T_n(a)\|_2 \to 0, \;\; \|B_nT_n(a)B_n - B_n\|_2 \to 0,\\ &\|(B_nT_n(a))^* - B_nT_n(a)\|_2 \to 0, \;\; \|(T_n(a)B_n)^* - T_n(a)B_n\|_2 \to 0 \;?\end{aligned}$$

Such a sequence $\{B_n\}$ is referred to as an asymptotic Moore-Penrose inverse of $T(a)$. In view of the (instable) conditions (15), the following result by Silbermann [25] is surprising: if $a \in PC$ and $T(a)$ is normally solvable, then $T(a)$ always has an asymptotic Moore-Penrose inverse. And what is the concern of this result with singular values ? One can easily show $T(a)$ has an asymptotic Moore-Penrose inverse if and only if there is a sequence $c_n \to 0$ and a number $d > 0$ such that

$$\Sigma(T_n(a)) \subset [0, c_n] \cup [d, \infty). \tag{18}$$

One says that $\Sigma(T_n(a))$ has the splitting property if (18) holds with $c_n \to 0$ and $d > 0$. Thus, Silbermann's result implies that if $a \in PC$ and $T(a)$ is normally solvable on l^2, then $\Sigma(T_n(a))$ has the splitting property.

Only recently, Roch and Silbermann [20], [21] were able to prove even much more. The sets $\Sigma(T_n(a))$ are said to have the k-splitting property, where $k \geq 0$ is an integer, if (18) is true for some sequence $c_n \to 0$ and some $d > 0$ and, in addition, exactly k singular values lie in $[0, c_n]$ and $n - k$ singular values are located in $[d, \infty)$

(here multiplicities are taken into account). Equivalently, $\Sigma(T_n(a))$ has the k-splitting property if and only if

$$\lim_{n\to\infty} s_k(T_n(a)) = 0 \quad \text{and} \quad \liminf_{n\to\infty} s_{k+1}(T_n(a)) > 0. \tag{19}$$

A normally solvable Toeplitz operator $T(a)$ on l^2 with a symbol $a \in PC$ is automatically Fredholm and therefore has some index $k \in \mathbf{Z}$. Roch and Silbermann [20], [21] discovered that then $\Sigma(T_n(a))$ has the $|k|$-splitting property. In other words, if $a \in PC$ and $T(a) \in \Phi_k(l^2)$ then (19) holds with k replaced by $|k|$. Notice that this Theorem 1.3 for $p = 2$.

In fact, it was the Roch and Silbermann papers [20], [21] which stimulated me to do some thinking about singular values. It was the feeling that the $|k|$-splitting property must have its root in the possibility of "ignoring $|k|$ dimensions" which led me to the observation that none of the works cited in this section makes use of the fact that $s_k(A_n)$ may alternatively be defined by (2), i.e. that singular values may also be viewed as approximation numbers. I then realized that some basic phenomena of [20] and [21] can be very easily understood by having recourse to (2) and that, moreover, using (2) is a good way to pass from l^2 and C^*-algebras to l^p and Banach algebras.

3. Toeplitz operators on l^p

We henceforth always assume that $1 < p < \infty$ and $1/p + 1/q = 1$.

Let M_p and $M_{\langle p\rangle}$ be as in Section 1. The set M_p can be shown to be a Banach algebra with pointwise algebraic operations and the norm $\|a\|_{M_p} := \|T(a)\|_p$. It is also well known that

$$M_p = M_q \subset M_2 = L^\infty$$

and

$$\|a\|_{M_p} = \|a\|_{M_q} \geq \|a\|_{M_2} = \|a\|_\infty \tag{20}$$

(see, e.g., [5, Section 2.5]). We remark that working with $M_{\langle p\rangle}$ instead of M_p is caused by the need of somehow reversing the estimate in (20). Suppose, for instance, $p > 2$ and $a \in M_{\langle p\rangle}$. Then $a \in M_{p+\varepsilon}$ for some $\varepsilon > 0$, and the Riesz-Thorin interpolation theorem gives

$$\|a\|_{M_p} \leq \|a\|_{M_2}^\gamma \, \|a\|_{p+\varepsilon}^{1-\gamma} = \|a\|_\infty^\gamma \, \|a\|_{M_{p+\varepsilon}}^{1-\gamma} \tag{21}$$

with some $\gamma \in (0,1)$ depending only on p and ε. The $\|a\|_{M_{p+\varepsilon}}$ on the right of (21) may in turn be estimated by $C_p(\|a\|_\infty + V_1(a))$ (recall Stechkin's inequality (5)) provided a has bounded total variation.

A bounded linear operator A on l^p is said to be normally solvable if its range, $\operatorname{Im} A$, is a closed subset of l^p. The operator A is called Fredholm if it is normally solvable and the spaces

$$\operatorname{Ker} A := \{x \in l^p : Ax = 0\} \quad \text{and} \quad \operatorname{Coker} A := l^p / \operatorname{Im} A$$

have finite dimensions. In that case the index $\operatorname{Ind} A$ is defined as

$$\operatorname{Ind} A := \dim \operatorname{Ker} A - \dim \operatorname{Coker} A.$$

We denote by $\Phi(l^p)$ the collection of all Fredholm operators on l^p and by $\Phi_k(l^p)$ the operators in $\Phi(l^p)$ whose index is k. The following four theorems are well known. Comments are at the end of this section.

Theorem 3.1. *Let $a \in M_p$.*

(a) *If a does not vanish identically, then the kernel of $T(a)$ on l^p or the kernel of $T(\overline{a})$ on l^q is trivial.*

(b) *The operator $T(a)$ is invertible on l^p if and only if $T(a) \in \Phi_0(l^p)$.*

Of course, part (b) is a simple consequence of part (a).

Theorem 3.2. *Let $a \in C \cap M_{\langle p \rangle}$. Then $T(a)$ is normally solvable on l^p if and only if $a(t) \neq 0$ for all $t \in \mathbf{T}$. In that case $T(a) \in \Phi(l^p)$ and*

$$\operatorname{Ind} T(a) = -\operatorname{wind} a,$$

where wind a *is the winding number of a about the origin.*

Now let $a \in PC$, $t \in \mathbf{T}$, and suppose $a(t-0) \neq a(t+0)$. We denote by

$$\mathcal{A}_p(a(t-0), a(t+0))$$

the circular arc at the points of which the line segment $[a(t-0), a(t+0)]$ is seen at the angle $\max\{2\pi/p, 2\pi/q\}$ and which lies on the right of the straight line passing first $a(t-0)$ and then $a(t+0)$ if $1 < p < 2$ and on the left of this line if $2 < p < \infty$. For $p = 2$, $\mathcal{A}_p(a(t-0), a(t+0))$ is nothing but the line segment $[a(t-0), a(t+0)]$ itself. Let $a_p^{\#}$ denote the closed, continuous, and naturally oriented curve which results from the (essential) range $\mathcal{R}(a)$ of a by filling in the arcs $\mathcal{A}_p(a(t-0), a(t+0))$ for each jump. In case this curve does not pass through the origin, we let $\operatorname{wind} a_p^{\#}$ be its winding number.

Theorem 3.3. *Let $a \in PC \cap M_{\langle p \rangle}$. Then $T(a)$ is normally solvable on l^p if and only if $0 \notin a_p^{\#}$. In that case $T(a) \in \Phi(l^p)$ and*

$$\operatorname{Ind} T(a) = -\operatorname{wind} a_p^{\#}.$$

For $a \in PC$ and $t \in \mathbf{T}$, put

$$\mathcal{O}_p\Big(a(t-0),\ a(t+0)\Big) := \bigcup_{r \in [p,q]} \mathcal{A}_r\Big(a(t-0),\ a(t+0)\Big). \tag{22}$$

If $a(t-0) \neq a(t+0)$ and $p \neq 2$, then $\mathcal{O}_p(a(t-0), a(t+0))$ is a certain lentiform set. Also for $a \in PC$, let

$$a_{[p,q]}^{\#} := \bigcup_{r \in [p,q]} a_r^{\#}.$$

Thus, $a^{\#}_{[p,q]}$ results from $\mathcal{R}(a)$ by filling in the sets (22) between the endpoints of the jumps. If $0 \notin a^{\#}_{[p,q]}$, then necessarily $0 \notin a^{\#}_2$ and we define $\operatorname{wind} a^{\#}_{[p,q]}$ as $\operatorname{wind} a^{\#}_2$ in this case.

From Theorem 3.3 we deduce that the conditions (i) to (iii) of Section 1 are equivalent to the following:

(i') $0 \notin a^{\#}_{[p,q]}$ and $\operatorname{wind} a^{\#}_{[p,q]} = k$;

(ii') $0 \in a^{\#}_p \cup a^{\#}_q$;

(iii') $0 \in a^{\#}_{[p,q]} \setminus (a^{\#}_p \cup a^{\#}_q)$.

For $a \in M_p$, let $T_n(a) \in \mathcal{B}(\mathbf{C}^n_p)$ be the operator given by the matrix (4). One says that the sequence $\{T_n(a)\} := \{T_n(a)\}_{n=1}^{\infty}$ is stable if

$$\limsup_{n\to\infty} \|T_n^{-1}(a)\|_p < \infty.$$

Here we follow the practice of putting

$$\|T_n^{-1}(a)\|_p = \infty \text{ if } T_n(a) \text{ is not invertible.}$$

In other words, $\{T_n(a)\}$ is stable if and only if $T_n(a)$ is invertible for all $n \geq n_0$ and there exists a constant $M < \infty$ such that $\|T_n^{-1}(a)\|_p \leq M$ for all $n \geq n_0$. From (3) we infer that

$$\{T_n(a)\} \text{ is stable } \iff \liminf_{n\to\infty} s_1^{(p)}(T_n(a)) > 0.$$

Theorem 3.4. (a) *If $a \in C \cap M_{\langle p\rangle}$ then*

$$\{T_n(a)\} \textit{ is stable } \iff 0 \notin a(\mathbf{T}) \textit{ and } \operatorname{wind} a = 0.$$

(b) If $a \in PC \cap M_{\langle p\rangle}$ then

$$\{T_n(a)\} \textit{ is stable } \iff 0 \notin a^{\#}_{[p,q]} \textit{ and } \operatorname{wind} a^{\#}_{[p,q]} = 0.$$

As already said, these theorems are well known. Theorem 3.1 is due to Coburn ($p = 2$) and Duduchava ($p \neq 2$), Theorem 3.2 is Gohberg and Feldman's, Theorem 3.3 is the result of many authors in the case $p = 2$ and was established by Duduchava for $p \neq 2$, Theorem 3.4 goes back to Gohberg and Feldman for $a \in C \cap M_{\langle p\rangle}$ (general p) and $a \in PC$ ($p = 2$), and it was obtained in the work of Verbitsky, Krupnik, Silbermann, and the author for $a \in PC \cap M_{\langle p\rangle}$ and $p \neq 2$. Precise historical remarks and full proofs are in [5].

Part (a) of Theorem 3.4 is clearly a special case of part (b). In fact, Theorem 3.4(b) may also be stated as follows: $\{T_n(a)\}$ contains a stable subsequence

$\{T_{n_j}(a)\}$ $(n_j \to \infty)$ if and only if $0 \notin a^{\#}_{[p,q]}$ and $\operatorname{wind} a^{\#}_{[p,q]} = 0$. Hence, we arrive at the conclusion that if $a \in PC \cap M_{\langle p \rangle}$, then

$$\begin{aligned} & s_1^{(p)}(T_n(a)) \to 0 \\ & \Longleftrightarrow \{T_n(a)\} \ \text{ is stable} \\ & \Longleftrightarrow 0 \in a^{\#}_{[p,q]} \ \text{ or } \ \Big(0 \notin a^{\#}_{[p,q]} \ \text{ and } \ \operatorname{wind} a^{\#}_{[p,q]} \neq 0\Big). \end{aligned}$$

At this point the question of whether the lowest approximation number of $T_n(a)$ goes to zero or not is completely disposed of for symbols $a \in PC \cap M_{\langle p \rangle}$.

4. Proof of Theorem 1.1.

Contrary to what we want, let us assume that there is a $c < \|T(a)\|_p$ such that $s^{(p)}_{n-k}(T_n(a)) \le c$ for all n in some infinite set $\mathcal{N}$. Since $s^{(p)}_{n-k}(T_n(a)) = \operatorname{dist}(T_n(a), \mathcal{F}^{(n)}_k)$, we can find $F_n \in \mathcal{F}^{(n)}_k$ $(n \in \mathcal{N})$ so that $\|T_n(a) - F_n\|_p \le c$. For $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$, we define

$$(x, y) := x_1 y_1 + \ldots + x_n y_n. \tag{23}$$

By [16, Lemma B.4.11], there exist $e^{(n)}_j \in \mathbf{C}^n_p$, $f^{(n)}_j \in \mathbf{C}^n_p$, $\gamma^{(n)}_j \in \mathbf{C}$ such that

$$F_n x = \sum_{j=1}^{k} \gamma^{(n)}_j \Big(x, f^{(n)}_j\Big) e^{(n)}_j \quad (x \in \mathbf{C}^n_p),$$

$\|e^{(n)}_j\|_p = 1$, $\|f^{(n)}_j\|_q = 1$, and

$$|\gamma^{(n)}_j| \le \|F_n\|_p \le \|T_n(a)\|_p + \|F_n - T_n(a)\|_p \le \|T(a)\|_p + c \tag{24}$$

for all $j \in \{1, \dots, k\}$.

Fix $x \in \mathbf{C}^n_p, y \in \mathbf{C}^n_q$ and suppose $\|x\|_p = 1$, $\|y\|_q = 1$. We then have

$$\left| \Big(T_n(a)x, y\Big) - \sum_{j=1}^{k} \gamma^{(n)}_j \Big(x, f^{(n)}_j\Big) \Big(e^{(n)}_j, y\Big) \right| \le \|T_n(a) - F_n\|_p \le c. \tag{25}$$

Clearly, $(T_n(a)x, y) \to (T(a)x, y)$. From (24) and the Bolzano-Weierstrass theorem we infer that the sequence $\{(\gamma^{(n)}_1, \dots, \gamma^{(n)}_k)\}_{n \in \mathcal{N}}$ has a converging subsequence. Without loss of generality suppose the sequence itself converges, i.e.

$$\Big(\gamma^{(n)}_1, \dots, \gamma^{(n)}_k\Big) \to (\gamma_1, \dots, \gamma_k) \in \mathbf{C}^k$$

as $n \in \mathcal{N}$ goes to infinity. The vectors $e^{(n)}_j$ and $f^{(n)}_j$ all belong to the unit sphere of l^p and l^q, respectively. Hence, by the Banach-Alaoglu theorem (see, e.g., [18, Theorem IV.21]), $\{e^{(n)}_j\}_{n \in \mathcal{N}}$ and $\{f^{(n)}_j\}_{n \in \mathcal{N}}$ have subsequences converging in the weak $*$-topology. Again we may without loss of generality assume that

$$e^{(n)}_j \to e_j \in l^p, \ f^{(n)}_j \to f_j \in l^q$$

in the weak $*$-topology as $n \in \mathcal{N}$ goes to infinity.

From (25) we now obtain that if $x \in l^p$ and $y \in l^q$ have finite support and $\|x\|_p = 1,\ \|y\|_q = 1$, then

$$\left|\Big(T(a)x, y\Big) - \sum_{j=1}^{k} \gamma_j (x, f_j)(e_j, y)\right| \le c.$$

This implies that

$$\|T(a) - F\|_p \le c \tag{26}$$

where F is the finite-rank operator given by

$$Fx := \sum_{j=1}^{k} \gamma_j (x, f_j) e_j \quad (x \in l^p). \tag{27}$$

Let $\|T(a)\|^{(\mathrm{ess})}$ denote the essential norm of $T(a)$ on l^p, i.e. the distance of $T(a)$ to the compact operators on l^p. By (26) and (27),

$$\|T(a)\|_p^{(\mathrm{ess})} \le \|T(a) - F\|_p \le c < \|T(a)\|_p.$$

However, one always has $\|T(a)\|_p^{(\mathrm{ess})} = \|T(a)\|_p$ (see, e.g., [5, Proposition 4.4(d)]). This contradiction completes the proof.

5. Proof of Theorem 1.2.

We will employ the following two results.

Theorem 5.1. *Let A be a bounded linear operator on l^p.*

(a) *The operator A is normally solvable on l^p if and only if*

$$k_A := \sup_{x \in l^p,\ \|x\|_p = 1} \operatorname{dist}(x, \operatorname{Ker} A) < \infty.$$

(b) *If M is a closed subspace of l^p and $\dim(l^p/M) < \infty$, then the normal solvability of $A|M : M \to l^p$ is equivalent to the normal solvability of $A : l^p \to l^p$.*

A proof is in [8, pp. 159–160]. □

Theorem 5.2. *If M is a k-dimensional subspace of $\mathbf{C}_p^n$, then there exists a projection $\Pi : \mathbf{C}_p^n \to \mathbf{C}_p^n$ such that $\operatorname{Im} \Pi = M$ and $\|\Pi\|_p \le k$.*

This is a special case of [16, Lemma B.4.9]. □

Theorem 1.2 is trivial in case a vanishes identically. So suppose $a \in M_p \setminus \{0\}$ and $T(a)$ is not normally solvable on l^p. Then the adjoint operator $T(\overline{a})$ is not normally solvable on l^q. By Theorem 3.1(a), $\operatorname{Ker} T(a) = \{0\}$ on l^p or $\operatorname{Ker} T(\overline{a}) = \{0\}$ on l^q.

Since $s_k^{(p)}(T_n(a)) = s_k^{(q)}(T_n(\overline{a}))$, we may a priori assume that $\operatorname{Ker} T(a) = \{0\}$ on l^p. Abbreviate $T(a)$ and $T_n(a)$ to A and A_n, respectively.

Define P_n on l^p by (13) and let

$$V := l^p \to l^p, \quad \{x_1, x_2, x_3, \ldots\} \mapsto \{0, x_1, x_2, x_3, \ldots\}.$$

As $A|\operatorname{Im} V^n : \operatorname{Im} V^n \to l^p$ has the same matrix as $AV^n : l^p \to l^p$, we deduce from Theorem 5.1(b) that there is no $n \geq 0$ such that AV^n is normally solvable. Note that $\operatorname{Ker}(AV^n) = \{0\}$ for all $n \geq 0$.

Let $l^p(n_1, n_2]$ denote the sequences $\{x_j\}_{j=1}^\infty \in l^p$ which are supported in $(n_1, n_2]$, i.e., for which $x_j = 0$ whenever $j \leq n_1$ or $j > n_2$.

Lemma 5.3. *There are $0 = N_0 < N_1 < N_2 < \ldots$ and $z_j \in l^p(N_{j-1}, N_j]$ $(j \geq 1)$ such that*

$$\|z_j\|_p = 1 \text{ and } \|Az_j\|_p \to 0 \text{ as } j \to \infty.$$

Proof. By Theorem 5.1(a), there is a $y_1 \in l^p$ such that $\|y_1\|_p = 2$ and $\|Ay_1\| < 1/2$. If N_1 is large enough, then $\|P_{N_1} y_1\|_p \geq 1$ and $\|AP_{N_1} y_1\|_p < 1$. Letting $z_1 := P_{N_1 y_1}/\|P_{N_1} y_1\|_p$ we get

$$z_1 \in l^p(0, N_1], \quad \|z_1\|_p = 1, \quad \|Az_1\|_p < 1.$$

Applying Theorem 5.1(a) to the operator AV^{N_1}, we see that there is an $y_2 \in l^p$ such that $\|y_2\|_p = 2$ and $\|AV^{N_1} y_2\|_p < 1/4$. For sufficiently large $N_2 > N_1$ we have $\|P_{N_2} V^{N_1} y_2\|_p \geq 1$ and $\|AP_{N_2} V^{N_1} y_2\|_p < 1/2$. Setting

$$z_2 := P_{N_2} V^{N_1} y_2 / \|P_{N_2} V^{N_1} y_2\|_p,$$

we therefore obtain

$$z_2 \in l^p(N_1, N_2], \quad \|z_2\|_p = 1, \quad \|Az_2\|_p < 1/2.$$

Continuing in this way we find z_j satisfying

$$z_j \in l^p(N_{j-1}, N_j], \quad \|z_j\|_p = 1, \quad \|Az_j\|_p < 1/j. \qquad \square$$

Contrary to the assertion of Theorem 1.2, let us assume that there exist $k \geq 1$ and $d > 0$ such that $s_k^{(p)}(A_n) \geq d$ for infinitely many n. We may without loss of generality assume that

$$s_k^{(p)}(A_n) \geq d \text{ for all } n \geq n_0. \tag{28}$$

Let $\varepsilon > 0$ be any number such that

$$2\varepsilon k^2 < d. \tag{29}$$

Choose z_j as in Lemma 5.3. Obviously, there are sufficiently large j and N such that

$$\|P_N z_l\|_p \geq 1/2, \quad \|AP_N z_l\|_p < \varepsilon \text{ for } l \in \{j+1, \ldots, j+k\}. \tag{30}$$

Since $P_N z_l \in l^p(N_{l-1,l}]$, it is clear that $P_N z_{j+1}, \ldots, P_N z_{j+k}$ are linearly independent. Now let $n \geq N$. By Theorem 5.2, there is a projection Π_n of $\mathbf{C}_p^n$ onto $\mathrm{span}\{P_N z_{j+1}, \ldots, P_N z_{j+k}\}$ for which $\|\Pi_n\|_p \leq k$. Let I_n stand for the identity operator on $\mathbf{C}_p^n$. The space $\mathrm{Im}\,(I_n - \Pi_n) = \mathrm{Ker}\,\Pi_n$ has the dimension $n-k$ and hence, $I_n - \Pi_n \in \mathcal{F}_{n-k}^{(n)}$. Every $x \in \mathbf{C}_p^n$ can be uniquely written in the form

$$x = \gamma_1 P_N z_{j+1} + \ldots + \gamma_k P_N z_{j+k} + w \quad \text{with} \quad w \in \mathrm{Ker}\,\Pi_n.$$

Thus,

$$\begin{aligned}
&\|A_n x - A_n(I_n - \Pi_n)x\|_p = \|A_n \Pi_n x\|_p \\
&= \|\gamma_1 A_n(P_N z_{j+1}) + \ldots + \gamma_k A_n(P_N z_{j+k})\|_p \leq |\gamma_1|\varepsilon + \ldots + |\gamma_k|\varepsilon,
\end{aligned} \tag{31}$$

the estimate resulting from (30). Taking into account that the sequences $P_N z_l$ have pairwise disjoint supports, we obtain from (30) that

$$\begin{aligned}
&\|\Pi_n x\|_p^p = \|\gamma_1 P_N z_{j+1} + \ldots + \gamma_k P_N z_{j+k}\|_p^p \\
&= |\gamma_1|^p \|P_N z_{j+1}\|_p^p + \ldots + |\gamma_k|^p \|P_N z_{j+k}\|_p^p \\
&\geq (1/2)^p \Big(|\gamma_1|^p + \ldots + |\gamma_k|^p\Big) \geq (1/2)^p \max_{1 \leq m \leq k} |\gamma_m|^p.
\end{aligned} \tag{32}$$

Combining (31) and (32) we get

$$\|A_n x - A_n(I_n - \Pi_n)x\|_p \leq \varepsilon k \max_{1 \leq m \leq k} |\gamma_m| \leq 2\varepsilon k \|\Pi_n x\|_p \leq 2\varepsilon k^2 \|x\|_p,$$

whence $s_k^{(p)}(A_n) = \mathrm{dist}\,(A_n, \mathcal{F}_{n-k}^{(n)}) \leq \|A_n - A_n(I - \Pi_n)\|_p \leq 2\varepsilon k^2$. By virtue of (29), this contradicts (28) and completes the proof.

6. Proof of Theorem 1.3.

The Hankel operator on l^p induced by a function $a \in M_p$ is given by the matrix

$$H(a) = (a_{j+k-1})_{j,k=1}^\infty.$$

For $a \in M_p$, define $\tilde{a} \in M_p$ by $\tilde{a}(e^{i\theta}) := a(e^{-i\theta})$. Clearly,

$$H(\tilde{a}) = (a_{-j-k+1})_{j,k=1}^\infty.$$

It is well known and easily seen that

$$T(ab) = T(a)T(b) + H(a)H(\tilde{b}) \tag{33}$$

for every $a, b \in M_p$. A finite section analogue of formula (33) reads

$$T_n(ab) = T_n(a)T_n(b) + P_n H(a)H(\tilde{b})P_n + W_n H(\tilde{a})H(b)W_n, \tag{34}$$

where P_n is as in (13) and W_n is defined by

$$W_n : \{x_1, x_2, x_3, \ldots\} \mapsto \{x_n, x_{n-1}, \ldots, x_1, 0, 0, \ldots\}.$$

The identity (34) first appeared in Widom's paper [31], a proof is also in [4, Proposition 3.6] and [5, Proposition 7.7].

We remark that $T_n(\tilde{a})$ is the transposed matrix of $T_n(a)$ and that the identity $T_n(\tilde{a}) = W_n T_n(a) W_n$ holds. In particular, we have

$$\begin{aligned} s_k^{(q)}(T_n(a)) &= \min\left\{ \|T_n(a) - F_{n-k}\|_q : F_{n-k} \in \mathcal{F}_{n-k}^{(n)} \right\} \\ &= \min\left\{ \|T_n(\tilde{a}) - G_{n-k}\|_p : G_{n-k} \in \mathcal{F}_{n-k}^{(n)} \right\} \\ &= \min\left\{ \|W_n(T_n(\tilde{a}) - G_{n-k})W_n\|_p : G_{n-k} \in \mathcal{F}_{n-k}^{(n)} \right\} \\ &= \min\left\{ \|T_n(a) - H_{n-k}\|_p : H_{n-k} \in \mathcal{F}_{n-k}^{(n)} \right\} \\ &= s_k^{(p)}(T_n(a)) \end{aligned} \tag{35}$$

(note also that W_n is an invertible isometry on $\mathbf{C}_p^n$).

To prove Theorem 1.3, we need the following two (well known) lemmas.

Lemma 6.1. *If $A, B, C \in \mathcal{B}(\mathbf{C}_p^n)$ then*

$$s_k^{(p)}(ABC) \le \|A\|_p \, s_k^{(p)}(B) \|C\|_p \ \textit{ for all } k.$$

This follows easily from the definition of $s_k^{(p)}$. □

Lemma 6.2. *If $b \in M_p$ and $\{T_n(b)\}$ is stable on l^p, then $T(b)$ is invertible on l^p and $T_n^{-1}(b)$ $(:= T_n^{-1}(b)P_n)$ converges strongly on l^p to $T^{-1}(b)$.*

This is obvious from the estimates

$$\|T_n^{-1}(b)P_n y - T^{-1}(b)y\|_p$$
$$\le \|T_n^{-1}(b)\|_p \, \|P_n y - T_n(b)P_n T^{-1}(b)y\|_p + \|P_n T^{-1}(b)y - T^{-1}(b)y\|_p,$$
$$\|x\|_p \le \liminf_{n\to\infty} \|T_n^{-1}(b)\|_p \, \|T(b)x\|_p, \quad \|\xi\|_q \le \liminf_{n\to\infty} \|T_n^{-1}(\tilde{b})\|_q \, \|T(\tilde{b})\xi\|_q. \quad \square$$

We now establish two propositions which easily imply Theorem 1.3.

Define χ_k by $\chi_k(e^{i\theta}) = e^{ik\theta}$. Using Theorem 3.1(b) and formula (33) one can readily see that if $a \in M_p$, then $T(a) \in \Phi_{-k}(l^p)$ if and only if $a = b\chi_k$ and $T(b)$ is invertible on l^p.

Propostion 6.3. *If $b \in M_p$ and $\{T_n(b)\}$ is stable on l^p then for every $k \in \mathbf{Z}$,*

$$\liminf_{n\to\infty} s_{|k|+1}^{(p)}\Big(T_n(b\chi_k)\Big) > 0.$$

Proof. We can assume that $k \ge 0$, since otherwise we may pass to adjoints. Because $\|T_n(\chi_{-k})\|_p = 1$, we obtain from Lemma 6.1 that

$$s_{k+1}^{(p)}\Big(T_n(b\chi_k)\Big) = s_{k+1}^{(p)}\Big(T_n(b\chi_k)\Big)\|T_n(\chi_{-k})\|_p$$
$$\ge s_{k+1}^{(p)}\Big(T_n(b\chi_k)T_n(\chi_{-k})\Big) = s_{k+1}^{(p)}\Big(T_n(b) - P_n H(b\chi_k)H(\chi_k)P_n\Big),$$

the latter equality resulting from (34) and the identities $H(\tilde{\chi}_{-k}) = H(\chi_k)$ and $H(\chi_{-k}) = 0$. As $\dim \operatorname{Im} H(\chi_k) = k$, we get that $F_k := P_n H(b\chi_k) H(\chi_k) P_n \in \mathcal{F}_k^{(n)}$, whence

$$\begin{aligned} s_{k+1}^{(p)}\Big(T_n(b) - F_k\Big) &= \inf\Big\{\|T_n(b) - F_k - G_{n-k-1}\|_p : G_{n-k-1} \in \mathcal{F}_{n-k-1}^{(n)}\Big\} \\ &\geq \inf\Big\{\|T_n(b) - H_{n-1}\|_p : H_{n-1} \in \mathcal{F}_{n-1}^{(n)}\Big\} = s_1^{(p)}(T_n(b)). \end{aligned}$$

Since $\{T_n(b)\}$ is stable, we infer from (3) that

$$\liminf_{n\to\infty} s_{k+1}^{(p)}(T_n(b\chi_k)) \geq \liminf_{n\to\infty} s_1^{(p)}(T_n(b)) > 0. \ \square$$

Proposition 6.4. *If $b \in M_p$ and $\{T_n(b)\}$ is stable on l^p then for every $k \in \mathbf{Z}$,*

$$\lim_{n\to\infty} s_{|k|}^{(p)}(T_n(b\chi_k)) = 0.$$

Proof. Again we may without loss of generality assume that $k \geq 0$. Using (34) and Lemma 6.1 we get

$$\begin{aligned} s_k^{(p)}(T_n(b\chi_k)) &= s_k^{(p)}\Big(T_n(\chi_k)T_n(b) + P_n H(\chi_k) H(\tilde{b}) P_n\Big) \\ &\leq \|T_n(b)\|_p \, s_k^{(p)}\Big(T_n(\chi_k) + P_n H(\chi_k) H(\tilde{b}) P_n T_n^{-1}(b)\Big). \end{aligned}$$

Put $A_n := T_n(\chi_k) + P_n H(\chi_k) H(\tilde{b}) P_n T_n^{-1}(b)$. We have

$$A_n = \begin{pmatrix} * & C_n \\ I_{n-k} & 0 \end{pmatrix} = \begin{pmatrix} * & 0 \\ I_{n-k} & 0 \end{pmatrix} + \begin{pmatrix} 0 & C_n \\ 0 & 0 \end{pmatrix} =: B_n + D_n,$$

the blocks being of size $k \times (n-k)$, $k \times k$, $(n-k) \times (n-k)$, $(n-k) \times k$, respectively. Clearly, B_n has rank $n-k$ and thus $B_n \in \mathcal{F}_{n-k}^{(n)}$. It follows that

$$s_k^{(p)}(A_n) = s_k^{(p)}(A_n - B_n) = s_k^{(p)}(D_n) \leq \|D_n\|_p = \|C_n\|_p,$$

and we are left with showing that $\|C_n\|_p \to 0$.

Let b_n $(n \in \mathbf{Z})$ be the Fourier coefficients of b, let $e_j \in l^p$ be the sequence whose only nonzero entry is a unit at the jth position, and recall the notation (23). We have $C_n = (c_{jl}^{(n)})_{j,l=1}^k$, and it is easily seen that $c_{jl}^{(n)}$ equals $(b_{-k+j-1}, \ldots, b_{-k+j-n})$ times the $(n-k+l)$th column of $T_n^{-1}(b)$:

$$c_{jl}^{(n)} = (b_{-k+j-1} \ldots b_{-k+j-n}) T_n^{-1}(b) P_n e_{n-k+l} = \Big(P_n f_{jk}, T_n^{-1}(b) P_n e_{n-k+l}\Big)$$

where

$$f_{jk} := \Big\{b_{-k+j-1},\ b_{-k+j-2},\ b_{-k+j-3},\ \ldots\Big\} = T(\chi_{-k+j-1}) T(\tilde{b}) e_1 \in l^q.$$

Consequently,

$$\begin{aligned} c_{jl}^{(n)} &= \Big(T_n^{-1}(\tilde{b}) P_n f_{jk},\ e_{n-k+l}\Big) \\ &= \Big(T^{-1}(\tilde{b}) f_{jk},\ e_{n-k+l}\Big) + \Big(T_n^{-1}(\tilde{b}) P_n f_{jk} - T^{-1}(\tilde{b}) f_{jk},\ e_{n-k+l}\Big). \end{aligned} \tag{36}$$

The first term on the right of (36) obviously converges to zero as $n \to \infty$. The second term of (36) is at most

$$\|T_n^{-1}(\tilde{b})P_n f_{jk} - T^{-1}(\tilde{b})f_{jk}\|_q \tag{37}$$

(note that $\|e_{n-k+l}\|_p = 1$). Our assumptions imply that $\{T_n(\tilde{b})\}$ is stable on l^q. We so deduce from Lemma 6.2 that (37) tends to zero as $n \to \infty$.

Thus, each entry of the $k \times k$ matrix C_n approaches zero as $n \to \infty$. This implies that $\|C_n\|_p \to 0$. □

Now let a be as in Theorem 1.3. Since $T(a) \in \Phi_{-k}(l^r)$ for all $r \in [p,q]$, we have $a = b\chi_k$ where $T(b) \in \Phi_0(l^r)$ for all $r \in [p,q]$. From Theorems 3.3 and 3.4(b) we conclude that $\{T_n(b)\}$ is stable on l^p. The assertions of Theorem 1.3 therefore follows from Propositions 6.3 and 6.4.

We remark that Propositions 6.3 and 6.4 actually yield more than Theorem 1.3. Namely, let Π_p^0 denote the collection of all symbols $b \in M_p$ for which $\{T_n(b)\}$ is stable on l^p and let Π_p be the set of all symbols $a \in M_p$ such that $a\chi_{-k} \in \Pi_p^0$ for some $k \in \mathbf{Z}$. Notice that

$$\Pi_p = \Pi_q \subset \bigcup_{r \in [p,q]} \Pi_r$$

and

$$G(C + H^\infty) \cup G(C + \overline{H^\infty}) \cup G(PQC) \subset \Pi_2 \neq L^\infty,$$

where $G(B)$ stands for the invertible elements of a unital Banach algebra B. The following corollary is immediate from Propositions 6.3 and 6.4.

Corollary 6.5. *If $a \in \Pi_p$ and $T(a) \in \Phi_k(l^p)$ then*

$$\Sigma^{(p)}(T_n(a)) := \left\{ s_1^{(p)}(T_n(a)), \ldots, s_n^{(p)}(T_n(a)) \right\}$$

has the $|k|$-splitting property. □

We also note that the proof of Proposition 6.4 gives estimates for the speed of convergence of $s_{|k|}^{(p)}(T_n(b\chi_k))$ to zero. For example, if $\sum_{n \in \mathbf{Z}} |n|^\mu |b_n| < \infty \quad (\mu > 0)$, then the finite section method is applicable to $T(b)$ on the space $l^{2,\mu}$ of all sequences $x = \{x_n\}_{n=1}^\infty$ such that

$$\|x\|_{2,\mu} := \left(\sum_{n=1}^\infty n^{2\mu} |x_n|^2 \right)^{1/2} < \infty$$

whenever $T(b)$ is invertible (see [17, pp. 106–107] or [5, Theorem 7.25]). Since

$$\|e_{n-k-l}\|_{2,-\mu} = (n - k + l)^{-\mu} = O(n^{-\mu}),$$

the proof of Proposition 6.4 implies the following result.

Corollary 6.6. *If $\sum_{n \in \mathbf{Z}} |n|^\mu |a_n| < \infty$ for some $\mu > 0$ and $T(a) \in \Phi_k(l^p)$ then*

$$s_{|k|}^{(p)}(T_n(a)) = O(n^{-\mu}) \quad \text{as} \quad n \to \infty. \ \square$$

7. Remarks on the Hilbert space case

Some aspects of the asymptotic behavior of the approximation numbers (= singular values) of matrices in $\mathcal{B}(\mathbf{C}_2^n)$ can be very easily understood by having recourse to the following well known fact (the "singular value decomposition").

Theorem 7.1. *If $A_n \in \mathcal{B}(\mathbf{C}_2^n)$ then there exist unitary matrices $U_n, V_n \in \mathcal{B}(\mathbf{C}_2^n)$ such that $A_n = U_n S_n V_n$ where*

$$S_n = \operatorname{diag}\Big(s_1(A_n), \dots, s_n(A_n)\Big). \ \square$$

Here and throughout this section we abbreviate $s_k^{(2)}(A_n)$ to $s_k(A_n)$.

To illustrate the usefulness of Theorem 7.1, we give another proof of Theorem 1.2 for $p = 2$. We still need the following result.

Theorem 7.2. *A bounded linear Hilbert space operator A is normally solvable if and only if there is a $d > 0$ such that*

$$\mathrm{sp}\,(A^*A) \subset \{0\} \cup [d, \infty).$$

For a proof see [10], [11], [20]. $\square$

Theorem 7.3. *Let $a \in L^\infty$ and suppose $T(a)$ is not normally solvable on l^2. Then $s_k(T_n(a)) \to 0$ as $n \to \infty$ for each $k \geq 1$.*

Proof. Assume there is a $k \geq 1$ such that $s_k(T_n(a))$ does not converge to zero. Let k_0 be the smallest k with this property. Then there are $n_1 < n_2 < \dots$ and $d > 0$ such that

$$s_{k_0}(T_{n_j}(a)) \geq d \ \text{ and } \ s_k(T_{n_j}(a)) \to 0 \ \text{ for } \ k < k_0. \tag{38}$$

To simplify notation, let us assume that $n_j = j$ for all j.

Write $T_n(a) = U_n S_n V_n$ as in Theorem 7.1. If $\lambda \notin \{0\} \cup [d^2, \infty)$, then (38) implies that $S_n^2 - \lambda I_n$ is invertible for all sufficiently large n, say for $n \geq n_0$, and that

$$\|(S_n^2 - \lambda I_n)^{-1}\|_2 \leq M(\lambda)$$

with some $M(\lambda) < \infty$ independent of n. Because

$$T_n^*(a)T_n(a) - \lambda I_n = V_n^*(S_n^2 - \lambda I_n)V_n,$$

it follows that $T_n^*(a)T_n(a) - \lambda I_n$ is invertible for $n \geq n_0$ and that

$$\|(T_n^*(a)T_n(a) - \lambda I_n)^{-1}\|_2 \leq M(\lambda).$$

Consequently, for every $x \in l^2$ we have

$$\|(T^*(a)T(a) - \lambda I)x\|_2 \geq \Big(1/M(\lambda)\Big)\|x\|_2,$$

which implies that $T^*(a)T(a) - \lambda I$ is invertible. Thus,

$$\mathrm{sp}\,\Big(T^*(a)T(a)\Big) \subset \{0\} \cup [d^2, \infty),$$

and Theorem 7.2 shows that $T(a)$ must be normally solvable. □

Things are more transparent by invoking a few (harmless) C^*-algebras. Let $\mathcal{S}$ denote the C^*-algebra of all sequences $\{A_n\} := \{A_n\}_{n=1}^{\infty}$ of operators $A_n \in \mathcal{B}(\mathbf{C}_2^n)$ such that

$$\|\{A_n\}\| := \sup_{n\geq 1} \|A_n\|_2 < \infty,$$

and let $\mathcal{S}_c$ be the C^*-algebra of all $\{A_n\} \in \mathcal{S}$ for which there exists a bounded linear operator A on l^2 such that $A_n \to A$ and $A_n^* \to A^*$ strongly. Finally, let $\mathcal{C}$ stand for the sequences $\{A_n\} \in \mathcal{S}$ for which $\|A_n\|_2 \to 0$. Clearly, $\mathcal{C}$ is a closed two-sided ideal in both $\mathcal{S}$ and $\mathcal{S}_c$.

Obviously, a sequence $\{A_n\} \in \mathcal{S}$ is stable if and only if $\{A_n\} + \mathcal{C}$ is invertible in $\mathcal{S}/\mathcal{C}$. Following [25] and [20], we call a sequence $\{A_n\} \in \mathcal{S}$ a Moore-Penrose sequence if there exists a sequence $\{B_n\} \in \mathcal{S}$ such that

$$\{A_nB_nA_n - A_n\} \in \mathcal{C}, \;\; \{B_nA_nB_n - B_n\} \in \mathcal{C}, \tag{39}$$

$$\Big\{(B_nA_n)^* - B_nA_n\Big\} \in \mathcal{C}, \;\; \Big\{(A_nB_n)^* - A_nB_n\Big\} \in \mathcal{C}. \tag{40}$$

An element a of a unital C^*-algebra $\mathcal{A}$ is said to be Moore-Penrose invertible if there is an element $a^+ \in \mathcal{A}$ such that

$$aa^+a = a, \;\; a^+aa^+ = a^+, \;\; (a^+a)^* = a^+a, \;\; (aa^+)^* = aa^+.$$

Thus, $\{A_n\} \in \mathcal{S}$ is a Moore-Penrose sequence if and only if $\{A_n\} + \mathcal{C}$ is Moore-Penrose invertible in $\mathcal{S}/\mathcal{C}$.

The following result is again from [10], [11], [20].

Theorem 7.4. *Let $\mathcal{A}$ be a unital C^*-algebra. An element $a \in \mathcal{A}$ is Moore-Penrose invertible in $\mathcal{A}$ if and only if there is a $d > 0$ such that* $\mathrm{sp}\,(a^*a) \subset \{0\} \cup [d, \infty)$. □

The next theorem is Roch and Silbermann's [20]. The proof given here is different from theirs.

Theorem 7.5. *A sequence $\{A_n\} \in \mathcal{S}$ is a Moore-Penrose sequence if and only if*

$$\Sigma(A_n) = \Big\{s_1(A_n), \ldots, s_n(A_n)\Big\}$$

has the splitting property.

Proof. Write $A_n = U_n S_n V_n$ as in Theorem 7.1. We have

$$\begin{aligned}&\|A_n B_n A_n - A_n\|_2 \to 0\\ \Longleftrightarrow\ &\|U_n S_n V_n B_n U_n S_n V_n - U_n S_n V_n\|_2 \to 0\\ \Longleftrightarrow\ &\|S_n (V_n B_n U_n) S_n - S_n\|_2 \to 0,\end{aligned}$$

and since analogous equivalences hold for the remaining three sequences in (39) and (40), we arrive at the conclusion that $\{A_n\}$ is a Moore-Penrose sequence if and only if $\{S_n\} + \mathcal{C}$ is Moore-Penrose invertible in $\mathcal{S}/\mathcal{C}$. By Theorem 7.4, this is equivalent to the condition

$$\operatorname{sp}_{\mathcal{S}/\mathcal{C}}\Big(\{S_n^2\} + \mathcal{C}\Big) \subset \{0\} \cup [d^2, \infty) \ \text{ for some } \ d > 0. \tag{41}$$

Let $\mathcal{D} \subset \mathcal{S}$ denote the sequences $\{A_n\}$ constituted by diagonal matrices A_n. From the elementary theory of C^*-algebras we get

$$\operatorname{sp}_{\mathcal{S}/\mathcal{C}}\Big(\{S_n^2\} + \mathcal{C}\Big) = \operatorname{sp}_{\mathcal{D}/(\mathcal{D}\cap\mathcal{C})}\Big(\{S_n^2\} + \mathcal{D} \cap \mathcal{C}\Big). \tag{42}$$

Consider the infinite diagonal matrix

$$\operatorname{diag}(S_1^2, S_2^2, \ldots) = \operatorname{diag}(\varrho_1, \varrho_2, \varrho_3, \ldots)$$

(here $\mathcal{S}_m \in \mathcal{B}(\mathbf{C}_2^m)$ and $\varrho_m \in \mathbf{C}$). Obviously, the spectrum on the right of (42) coincides with the set $\mathcal{P}\{\varrho_m\}$ of the partial limits of the sequence $\{\varrho_m\}$. Consequently, (41) holds if and only if

$$\mathcal{P}\{\varrho_m\} \subset \{0\} \cup [d^2, \infty) \ \text{ for some } \ d > 0,$$

which is easily seen to be equivalent to the splitting property of $\Sigma(A_n)$. □

Also as in [20], we call a sequence $\{A_n\} \in \mathcal{S}$ an exact Moore-Penrose sequence if $\{A_n^+\}$ belongs to $\mathcal{S}$; here $A_n^+ \in \mathcal{B}(\mathbf{C}_2^n)$ is the Moore-Penrose inverse of A_n.

Proposition 7.6. *Let $\{A_n\}$ be a sequence in $\mathcal{S}_c$ and let A be the strong limit of A_n. Then the following are equivalent:*

(i) *A_n^+ is strongly convergent;*

(ii) *A is normally solvable and $A_n^+ \to A^+$ strongly;*

(iii) *A is normally solvable and $\{A_n\}$ is an exact Moore-Penrose sequence.*

The simple proof is omitted. □

The following theorem was by means of different methods established in [20].

Theorem 7.7. *A sequence $\{A_n\} \in \mathcal{S}$ is an exact Moore-Penrose sequence if and only if there is a $d > 0$ such that*

$$\Sigma(A_n) \subset \{0\} \cup [d, \infty) \ \textit{ for all } \ n \geq 1. \tag{43}$$

Proof. As in the proof of Theorem 7.5 we see that $\{A_n\}$ is an exact Moore-Penrose sequence if and only if $\{S_n\}$ enjoys this property. Write

$$\operatorname{diag}(S_1, S_2, \ldots) = \operatorname{diag}(\mu_1, \mu_2, \mu_3, \ldots)$$

(where again $S_m \in \mathcal{B}(\mathbf{C}_2^m)$ and $\mu_n \in \mathbf{C}$) and define $f : [0,\infty) \to [0,\infty)$ by

$$f(x) := \begin{cases} x^{-1} & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0. \end{cases}$$

Since

$$\operatorname{diag}(S_1^+, S_2^+, \ldots) = \operatorname{diag}\Big(f(\mu_1), f(\mu_2), f(\mu_3), \ldots\Big),$$

we conclude that $\{S_n^+\} \in \mathcal{S}$ if and only if $\{f(\mu_m)\}$ is a bounded sequence, which is equivalent to (43). □

Now let $A_n = T_n(a)$ with $a \in L^\infty$. If $\{T_n(a)\}$ is a Moore-Penrose sequence, then $T(a)$ must obviously be normally solvable. Thus, from Theorem 3.3 (for $p = 2$) and Theorem 1.3 (for $p = 2$) we deduce that if $a \in PC$, then $\{T_n(a)\}$ is a Moore-Penrose sequence if and only if $T(a)$ is Fredholm.

The following result, which is also taken from [20], characterizes the exact Moore-Penrose sequences constituted by the truncations of an infinite Toeplitz matrix. Our proof is again different from the one of [20].

Theorem 7.8. *Let $a \in PC$. Then $\{T_n(a)\}$ is an exact Moore-Penrose sequence if and only if $T(a)$ is Fredholm and*

$$\dim \operatorname{Ker} T_n(a) = |\operatorname{Ind} T(a)| \tag{44}$$

for all sufficiently large n.

Proof. If $\{T_n(a)\}$ is an exact Moore-Penrose sequence, then $T(a)$ is normally solvable and thus Fredholm. Let $T(a) \in \Phi_k(l^2)$. Then

$$s_{|k|}(T_n(a)) \to 0 \quad \text{and} \quad s_{|k|+1}(T_n(a)) \ge d > 0$$

by virtue of Theorem 1.3 (for $p = 2$). Since

$$\operatorname{dist}\Big(T_n(a),\ \mathcal{F}_{n-|k|-1}^{(n)}\Big) > 0,$$

we see that

$$\operatorname{rank} T_n(a) \ge n - |k|. \tag{45}$$

From Theorem 7.7 we deduce that $\{T_n(a)\}$ is an exact Moore-Penrose sequence if and only if $s_{|k|}(T_n(a)) = 0$ for all $n \ge n_0$. Because

$$s_{|k|}(T_n(a)) = \operatorname{dist}\Big(T_n(a),\ \mathcal{F}_{n-|k|}^{(n)}\Big)$$

and $\mathcal{F}^{(n)}_{n-|k|}$ is a closed subset of $\mathcal{B}(\mathbf{C}_2^n)$, we have $s_{|k|}(T_n(a)) = 0$ if and only if

$$\operatorname{rank} T_n(a) \le n - |k|. \tag{46}$$

Combining (45) and (46) we obtain that $\{T_n(a)\}$ is an exact Moore-Penrose sequence if and only if $T(a) \in \Phi_k(l^2)$ for some $k \in \mathbf{Z}$ and

$$\dim \operatorname{Ker} T_n(a) = n - \operatorname{rank} T_n(a) = |k|$$

for all $n \ge n_0$. □

8. The Heing-Hellinger theorem

Of course, condition (44) is difficult to check. In this section we give a new proof of the Heinig-Hellinger theorem, which provides a criterion (in terms of only the symbol a) for (44) to hold.

If $a \in PC$ and $T(a)$ is Fredholm of index zero and thus invertible, then the sequence $\{T_n(a)\}$ is stable (Theorems 3.3 and 3.4 for $p = 2$). In this case $\Sigma(T_n(a)) \subset [d, \infty)$ and $\dim \operatorname{Ker} T_n(a) = 0$ for all sufficiently large n and hence each of Theorems 7.7 and 7.8 yields that $\{T_n(a)\}$ is an exact Moore-Penrose sequence; however, we have $T_n^+(a) = T_n^{-1}(a)$ for all sufficiently large n and therefore consideration of Moore-Penrose inverses is not at all necessary in this situation.

The really interesting case is the one in which $T(a)$ is Fredholm of nonzero index. The rest of this section is devoted to the proof of the following result.

Theorem 8.1 (Heinig and Hellinger). *Let $a \in PC$. Suppose $T(a)$ is Fredholm on l^2 and $\operatorname{Ind} T(a) \ne 0$. If $\operatorname{Ind} T(a) < 0$, then the following are equivalent:*

(i) $\dim \operatorname{Ker} T_n(a) = |\operatorname{Ind} T(a)|$ *for all sufficiently large n;*

(ii) $\operatorname{Ker} T(\tilde{a}) \subset \operatorname{Im} P_{n_0}$ *for some $n_0 \ge 1$;*

(iii) *the Fourier coefficients $(a^{-1})_{-m}$ are zero for all sufficiently large m.*

If $\operatorname{Ind} T(a) > 0$, *then the following are equivalent:*

(i') $\dim \operatorname{Ker} T_n(a) = \operatorname{Ind} T(a)$ *for all sufficiently large n;*

(ii') $\operatorname{Ker} T(a) \subset \operatorname{Im} P_{n_0}$ *for some $n_0 \ge 1$;*

(iii') $(a^{-1})_m = 0$ *for all sufficiently large m.*

For the sake of definiteness, let us assume that $\operatorname{Ind} T(a) = -k < 0$. The proofs of the implications (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) are easy.

Proof of the implication (iii) $\Rightarrow$ (ii). Let $x \in \operatorname{Ker} T(\tilde{a})$. Then, by (33),

$$T(\tilde{a}^{-1})T(\tilde{a}) = I - H(\tilde{a}^{-1})H(a),$$

which shows that $x = H(\tilde{a}^{-1})H(a)x$, and since $H(\tilde{a}^{-1})$ has only a finite number of nonzero rows, it follows that $x_m = 0$ for all sufficiently large m. □

Proof of the implication (ii) $\Rightarrow$ (i). If n is large enough then $s_{k+1}(T_n(\tilde{a})) \geq d > 0$ by Theorem 1.3 (or Proposition 6.3), whence $\operatorname{rank} T_n(\tilde{a}) > n - k + 1$ and thus,

$$\dim \operatorname{Ker} T_n(\tilde{a}) < k + 1. \tag{47}$$

If $x \in \operatorname{Ker} T(\tilde{a}) \subset \operatorname{Im} P_{n_0}$ and $n \geq n_0$, then $T_n(\tilde{a})P_n x = P_n T(\tilde{a})x = 0$, which implies that

$$\dim \operatorname{Ker} T_n(\tilde{a}) \geq \dim \operatorname{Ker} T(\tilde{a}) = k \tag{48}$$

(recall Theorem 3.1(a) for the last equality). Clearly, equality (i) follows from (47) and (48). □

The proof of the implication (i) $\Rightarrow$ (iii) is less trivial and is based on the following deep theorem. Recall that χ_n is defined by $\chi_n(t) = t^n$ for $t \in \mathbf{T}$.

Theorem 8.2 (Heinig). *Let $a \in L^\infty$ and let $k > 0$ be an integer. Then*

$$\dim \operatorname{Ker} T_n(a) = k \quad \textit{for all sufficiently large } n$$

if and only if a or $\tilde{a}$ is of the form $\chi_{p+k}(r+h)$ where h is a function in H^∞, r is a rational function in L^∞, r has exactly p poles in the open unit disk $\mathbf{D}$ (multiplicities taken into account), r has no pole at the origin, and $r(0) + h(0) \neq 0$.

A proof is in [12, Satz 6.2 and formula (8.4)]. Also see [14, Theorem 8.6]. □

Proof of the implication (i) $\Rightarrow$ (iii). Let $\chi_{p+k}(r+h)$ be the representation of a or $\tilde{a}$ ensured by Theorem 8.2 and put $b := \chi_{p+k}(r+h)$. Denote by $\alpha_1, \ldots, \alpha_p$ and $\beta_1, \ldots, \beta_q$ the poles of r inside and outside $\mathbf{T}$, respectively. For $t \in \mathbf{T}$,

$$\begin{aligned} r(t) &= \frac{u_+(t)}{(t-\alpha_1)\ldots(t-\alpha_p)(t-\beta_1)\ldots(t-\beta_q)} \\ &= \frac{t^{-p}v_+(t)}{(1-\alpha_1/t)\ldots(1-\alpha_p/t)(1-t/\beta_1)\ldots(1-t/\beta_q)} \end{aligned}$$

with polynomials $u_+, v_+ \in H^\infty$. Clearly,

$$s_+(t) := (1-t/\beta_1)^{-1} \ldots (1-t/\beta_q)^{-1} \in H^\infty.$$

Letting

$$c_+(t) := t^k v_+(t) s_+(t) + t^{p+k}(1-\alpha_1/t)\ldots(1-\alpha_p/t)h(t),$$

we get

$$b(t) = (1-\alpha_1/t)^{-1} \ldots (1-\alpha_p/t)^{-1} c_+(t).$$

The function c_+ lies in H^∞ and has a zero of order at least k at the origin. Obviously, $(1-\alpha_1/t)^{-1} \ldots (1-\alpha_p/t)^{-1}$ is a function which together with its inverse belongs to

$\overline{H^\infty}$. If c_+ would have infinitely many zeros in $\mathbf{D}$, then $T(c_+)$ and thus $T(b)$ were not Fredholm (see, e.g., [5, Theorem 2.64]). Hence, c_+ has only a finite number $\lambda \geq k$ of zeros in $\mathbf{D}$. It follows that $\operatorname{Ind} T(c_+) = -\lambda$ (again see, e.g., [5, Theorem 2.64]) and therefore $\operatorname{Ind} T(b) = \operatorname{Ind} T(c_+) = -\lambda$. If $b = a$, then λ must equal k. Consequently, $c_+(z) = z^k \varphi_+(z)$ with φ_+ and φ_+^{-1} in H^∞. This implies that

$$a^{-1}(t) = t^{-k}(1-\alpha_1/t)\dots(1-\alpha_p/t)\varphi_+^{-1}(t)$$

has only finitely many nonzero Fourier coefficients with negative index. If b would be equal to $\tilde{a}$, it would result that $\operatorname{Ind} T(\tilde{a})$ is negative, which is impossible due to the equality $\operatorname{Ind} T(\tilde{a}) = -\operatorname{Ind} T(a)$. □

Corollary 8.3. *If $a \in PC \setminus C$ then $\{T_n(a)\}$ is an exact Moore-Penrose sequence on l^2 if and only if $\{T_n(a)\}$ is stable on l^2.*

Proof. The "if part" is trivial. To prove the "only if" portion, suppose $\{T_n(a)\}$ is an exact Moore-Penrose sequence. Then $T(a)$ is Fredholm by Theorem 7.8. If $T(a)$ has index zero, then $\{T_n(a)\}$ is stable. If $\operatorname{Ind} T(a) \neq 0$, then Theorem 7.8 and the implication (i) $\Rightarrow$ (iii) of Theorem 8.1 tell us that a^{-1} is a polynomial times a function in H^∞ or $\overline{H^\infty}$. As functions in H^∞ or $\overline{H^\infty}$ cannot have jumps, this case is impossible. □

We remark that Heinig and Hellinger [13] proved the equivalence (i) $\Leftrightarrow$ (iii) of Theorem 8.2 for symbols in the Wiener algebra W. Corollary 8.3 was known to Silbermann and led him to the introduction of condition (ii). In the case of block Toeplitz matrices, (iii) and (ii) are no longer equivalent; Silbermann proved that then the validity of (15) for some $n_0 \geq 1$ implies that

$$\{T_n(a)\} \text{ is an exact Moore-Penrose sequence,} \tag{49}$$

and he conjectures that (49) is even equivalent to (15) for some $n_0 \geq 1$ (see [25]). The proofs of [13] and [25] differ from the proof given above.

9. l^p versus l^2

As shown in the previous section, many l^2 results can be derived with the help of Theorem 7.1, which reduces problems for $\{A_n\}$ to questions about the infinite diagonal operator

$$\operatorname{diag}\left(s_1^{(2)}(A_1),\, s_1^{(2)}(A_2),\, s_2^{(2)}(A_2),\, s_1^{(2)}(A_3),\, s_2^{(2)}(A_3),\, s_3^{(2)}(A_3),\, \dots\right).$$

It would therefore be very nice to have an analogous result for l^p. For example, one could ask the following: given $A_n \in \mathcal{B}(\mathbf{C}_p^n)$, are there invertible isometries $U_n, V_n \in \mathcal{B}(\mathbf{C}_p^n)$ and a diagonal matrix $S_n \in \mathcal{B}(\mathbf{C}_p^n)$ such that $A_n = U_n S_n V_n$? If the answer were "yes", we had

$$\Sigma^{(p)}(A_n) = \Sigma^{(p)}(S_n),$$

and Theorem 11.11.3 of [16] would tell us that $\Sigma^{(p)}(S_n)$ is the collection of the moduli of the diagonal elements of S_n.

However, the answer to the above question is "no". The reason is the dramatic loss of symmetry when passing from l^2 to l^p. Looking at the (real) unit spheres

$$\mathbf{S}_1^{(p)} := \Big\{(x,y) \in \mathbf{R}^2 : |x|^p + |y|^p = 1\Big\},$$

we see that $\mathbf{S}_1^{(2)}$ has the symmetry group $O(2)$, while the symmetry group of $\mathbf{S}_1^{(p)}$ $(p \neq 2)$ is the dieder group D_4, which contains only 8 elements. Equivalently, the invertible isometries in $\mathcal{B}(\mathbf{C}_2^2)$ are the 2×2 unitary matrices, whereas a matrix $U_2 \in \mathcal{B}(\mathbf{C}_p^2)$ $(p \neq 2)$ is an invertible isometry if and only if

$$U_2 = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} \text{ or } U_2 = \begin{pmatrix} 0 & \lambda \\ \mu & 0 \end{pmatrix} \text{ with } (\lambda,\mu) \in \mathbf{T}^2.$$

Thus, a matrix $A_2 \in \mathcal{B}(\mathbf{C}_p^2)$ $(p \neq 2)$ is of the form $A_2 = U_2 S_2 V_2$ with invertible isometries U_2, V_2 and a diagonal matrix S_2 if and only if

$$A_2 = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \text{ or } A_2 = \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix} \text{ with } (a,b) \in \mathbf{C}^2.$$

I even suspect that relaxing the above question will not be successful.

Conjecture 9.1. *Fix $p \neq 2$ and let $1/p + 1/q = 1$. There is no number $M \in (1,\infty)$ with the following property: given any sequence $\{A_n\}$ of matrices $A_n \in \mathcal{B}(\mathbf{C}_p^n)$ such that $\sup \|A_n\|_p < \infty$ and $\sup \|A_n\|_q < \infty$, there are invertible matrices $U_n, V_n \in \mathcal{B}(\mathbf{C}_p^n)$ and diagonal matrices $S_n \in \mathcal{B}(\mathbf{C}_p^n)$ such that $A_n = U_n S_n V_n$ and*

$$\|U_n\|_p \le M, \ \ \|U_n^{-1}\|_p \le M, \ \ \|V_n\|_p \le M, \ \ \|V_n^{-1}\|_p \le M$$

for all n.

Finally, for the reader's convenience, we add a proof of (3).

Proposition 9.2. *If $A \in \mathcal{B}(\mathbf{C}_p^n)$, then $s_1^{(p)}(A) = 1/\|A^{-1}\|_p$ if A is invertible and $s_1^{(p)}(A) = 0$ if A is not invertible.*

Proof. Suppose A is not invertible. Then $\operatorname{Ker} A \neq \{0\}$. Let Z be any direct complement of $\operatorname{Ker} A$ in $\mathbf{C}_p^n$ and let $P : \mathbf{C}_p^n \to Z$ be the projection onto Z parallel to $\operatorname{Ker} A$. Clearly, $P \in \mathcal{F}_{n-1}^{(n)}$ and thus $F := AP \in \mathcal{F}_{n-1}^{(n)}$. If $x \in \mathbf{C}^n$, then $x = x_0 + x_1$ with $x_0 \in \operatorname{Ker} A$ and $x_1 = Px \in Z$. Therefore

$$(A - F)x = Ax - APx = A(x_0 + Px) - APx = 0,$$

which implies that $A - F = 0$ and hence $\operatorname{dist}(A, \mathcal{F}_{n-1}^{(n)}) = 0$.

Now suppose A is invertible. We then have

$$\|A^{-1}\|_p = \sup_{x\neq 0} \frac{\|A^{-1}x\|_p}{\|x\|_p} = \sup_{z \neq 0} \frac{\|z\|_p}{\|Az\|_p} = \left(\inf_{z\neq 0} \frac{\|Az\|_p}{\|z\|_p}\right)^{-1},$$

whence

$$1/\|A^{-1}\|_p = \inf_{z\neq 0} \frac{\|Az\|_p}{\|z\|_p} = \min_{\|z\|_p=1} \|Az\|_p =: \|Ae_0\|_p \tag{50}$$

with some $e_0 \in \mathbf{C}_p^n$ of norm 1. Put $\operatorname{span}\{e_0\} = \{\lambda e_0 : \lambda \in \mathbf{C}\}$ and let X be any direct complement of $\operatorname{span}\{e_0\}$ in $\mathbf{C}_p^n$. The functional

$$\varphi : \operatorname{span}\{e_0\} \to \mathbf{C}, \quad \lambda e_0 \mapsto \lambda$$

clearly has the norm 1. By the Hahn-Banach theorem, there is a functional $\Phi : \mathbf{C}_p^n \to \mathbf{C}$ such that $\Phi(\lambda e_0) = \lambda$ and $\|\Phi\| = 1$. Define $F \in \mathcal{B}(\mathbf{C}_p^n)$ by $Fx := Ax - \Phi(x)Ae_0$. Since

$$F(\lambda e_0) = \lambda Ae_0 - \lambda Ae_0 = 0,$$

we see that $F \in \mathcal{F}_{n-1}^{(n)}$. Because

$$\|Ax - Fx\|_p = \|\Phi(x)Ae_0\|_p = |\Phi(x)|\,\|Ae_0\|_p \leq \|x\|_p\,\|Ae_0\|_p,$$

it results that $\|A - F\|_p \leq \|Ae_0\|_p$. From (50) we therefore deduce that $s_1^{(p)}(A) \leq 1/\|A^{-1}\|_p$.

To prove that $s_1^{(p)}(A) \geq 1/\|A^{-1}\|_p$, let G be any matrix in $\mathcal{F}_{n-1}^{(n)}$. If $\|I - A^{-1}G\|_p$ were less than 1, then $A^{-1}G$ and thus G were invertible, which is impossible. Thus $\|I - A^{-1}G\|_p \geq 1$. We therefore have

$$1 \leq \|I - A^{-1}G\|_p = \|A^{-1}(A - G)\|_p \leq \|A^{-1}\|_p\,\|A - G\|_p,$$

which implies that $1/\|A^{-1}\|_p \leq \|A - G\|_p$. As $G \in \mathcal{F}_{n-1}^{(n)}$ was arbitrary, it follows that $1/\|A^{-1}\|_p \leq s_1^{(p)}(A)$. □

References

[1] F. Avram: On bilinear forms in Gaussian random variables and Toeplitz matrices. *Probab. Theory Related Fields* 79 (1988), 37–45.

[2] A. Böttcher, S. Grudsky: Toeplitz operators with discontinuous symbols: phenomena beyond piecewise continuity. *Operator Theory: Advances and Applications* 90 (1996) 55–118.

[3] A. Böttcher, S. Silbermann: Über das Reduktionsverfahren für diskrete Wiener-Hopf Gleichungen mit unstetigem Symbol. *Z. Analysis Anw.* 1 (1982), 1–5.

[4] A. Böttcher, B. Silbermann: *Invertibility and Asymptotics of Toeplitz Matrices.* Akademie-Verlag, Berlin 1983.

[5] A. Böttcher, B. Silbermann: *Analysis of Toeplitz Operators.* Akademie-Verlag, Berlin 1989 and Springer-Verlag, Berlin, Heidelberg, New York 1990.

[6] K.M. Day: Measures associated with Toeplitz matrices generated by Laurent expansions of an arbitrary rational function. *Trans. Amer. Math. Soc.* 209 (1975), 175–183.

[7] I. Gohberg, I.A. Feldman: *Convolution Equations and Projection Methods for Their Solution.* Amer. Math. Soc., Providence, RI, 1974 [Russian original: Nauka, Moscow 1971; German transl.: Akademie-Verlag, Berlin 1974].

[8] I. Gohberg, N. Krupnik: *One-Dimensional Linear Singular Integral Equations.* Vol. I, Birkhäuser Verlag, Basel, Boston, Berlin 1992 [Russian original: Shtiintsa, Kishinev 1973; German transl.: Birkhäuser Verlag, Basel 1979].

[9] U. Grenander, G. Szegö: *Toeplitz Forms and Their Applications.* Univ. of California Press, Berkeley 1958 [Russian transl.: Izd. Inostr. Lit., Moscow 1961].

[10] R. Harte, M. Mbekhta: On generalized inverses in C^*-algebras. *Studia Math.* 103 (1992), 71–77.

[11] R. Harte, M. Mbekhta: Generalized inverses in C^*-algebras. II. *Studia Math.* 106 (1993), 129–138.

[12] G. Heinig: Endliche Toeplitzmatrizen und zweidimensionale diskrete Wiener-Hopf-Operatoren mit homogenem Symbol. *Math. Nachr.* 82 (1978), 29–68.

[13] G. Heinig, F. Hellinger: The finite section method for Moore-Penrose inversion of Toeplitz operators. *Integral Equations and Operator Theory* 19 (1994), 419–446.

[14] G. Heinig, K. Rost: *Algebraic Methods for Toeplitz-Like Matrices and Operators.* Akademie-Verlag, Berlin 1984 and Birkhäuser Verlag, Basel, Boston, Stuttgart 1984.

[15] S.V. Parter: On the distribution of the singular values of Toeplitz matrices. *Lin. Algebra Appl.* 80 (1986), 115–130.

[16] A. Pietsch: *Operator Ideals.* Deutscher Verlag d. Wiss., Berlin 1978 [Russian transl.: Mir, Moscow 1982].

[17] S. Prössdorf, B. Silbermann: *Projektionsverfahren und die näherungsweise Lösung singulärer Gleichungen.* Teubner, Leipzig 1977.

[18] M. Reed, B. Simon: *Methods of Modern Mathematical Physics. Vol. 1: Functional Analysis.* Academic Press, New York, London 1972 [Russian transl.: Mir, Moscow 1977].

[19] S. Roch, B. Silbermann: Limiting sets of eigenvalues and singular values of Toeplitz matrices. *Asymptotic Analysis* 8 (1994), 293–309.

[20] S. Roch, B. Silbermann: Index calculus for approximation methods and singular value decomposition. Preprint, TU Chemnitz-Zwickau 1996.

[21] S. Roch, B. Silbermann: A note on singular values of Cauchy-Toeplitz matrices. Preprint TU Chemnitz-Zwickau 1996.

[22] P. Schmidt, F. Spitzer: The Toeplitz matrices of an arbitrary Laurent polynomial. *Math. Scand.* 8 (1960), 15–38.

[23] B. SILBERMANN: Lokale Theorie des Reduktionsverfahrens für Toeplitzoperatoren. *Math. Nachr.* 104 (1981), 137–146.

[24] B. SILBERMANN: On the limiting set of singular values of Toeplitz matrices. *Lin. Algebra Appl.* 182 (1993), 35–43.

[25] B. SILBERMANN: Asymptotic Moore-Penrose inversion of Toeplitz operators. *Lin. Algebra Appl.*, to appear 1996.

[26] S.TREIL: Invertibility of Toeplitz operators does not imply applicability of the finite section method. *Dokl. Akad. Nauk SSSR* 292 (1987), 563–567 [Russian].

[27] E.E. TYRTYSHNIKOV: Singular values of Cauchy-Toeplitz matrices. *Lin. Algebra Appl.* 161 (1992), 99–116.

[28] E.E. TYRTYSHNIKOV: New theorems on the distribution of eigenvalues and singular values of multilevel Toeplitz matrices. *Dokl. Adad. Nauk* 333 (1993), 300–303 [Russian].

[29] E.E. TYRTYSHNIKOV: A unifying appraoch to some old and new theorems on distribution and clustering. *Lin. Algebra Appl.* 232 (1996), 1–43.

[30] I.E. VERBITSKY, N. KRUPNIK: On the applicability of the finite section method to discrete Wiener-Hopf equations with piecewise continuous symbol. *Matem. Issled.* 45 (1977), 17–28 [Russian].

[31] H. WIDOM: Asymptotic behavior of block Toeplitz matrices and determinants. II. *Adv. Math* 21 (1976), 1–29.

[32] H. WIDOM: On the singular values of Toeplitz matrices. *Z. Analysis Anw.* 8 (1989), 221–229.

Albrecht Böttcher
Fakultät für Mathematik
TU Chemnitz-Zwickau
09107 Chemnitz
Germany

On the Finiteness of Ш for Motives Associated to Modular Forms

Amnon Besser[1]

Received: February 1, 1996
Revised: January 15, 1997

Communicated by Don Blasius

Abstract. Let f be a modular form of even weight on $\Gamma_0(N)$ with associated motive $\mathcal{M}_f$. Let K be a quadratic imaginary field satisfying certain standard conditions. We improve a result of Nekovář and prove that if a rational prime p is outside a finite set of primes depending only on the form f, and if the image of the Heegner cycle associated with K in the p-adic intermediate Jacobian of $\mathcal{M}_f$ is not divisible by p, then the p-part of the Tate-Šafarevič group of $\mathcal{M}_f$ over K is trivial. An important ingredient of this work is an analysis of the behavior of "Kolyvagin test classes" at primes dividing the level N. In addition, certain complications, due to the possibility of f having a Galois conjugate self-twist, have to be dealt with.

1991 Mathematics Subject Classification: 11G18, 11F66, 11R34, 14C15.

1 Introduction

Let f be a new form of even weight $2r$ for the group $\Gamma_0(N)$, let $\mathcal{M}_f$ be the r-th Tate twist of the motive associated to f by Jannsen [Jan88b] and Scholl [Sch90]. For all but a finite number of primes p there is a canonical choice of free $\mathbb{Z}_p$-lattice $T_p(\mathcal{M}_f)$ with a continuous action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ such that $T_p(\mathcal{M}_f)\otimes\mathbb{Q}$ is the p-adic realization of $\mathcal{M}_f$. In [Nek92], Nekovář showed that under certain assumption one could apply the Kolyvagin method of Euler systems to $\mathcal{M}_f$ and obtained, among other things, the following result:

Theorem 1.1. *Let K be a quadratic imaginary field of discriminant D in which all primes dividing N split, and let p be a prime not dividing $2N$. Let $T_p(\mathcal{M}_f)$ be the p-adic realization of $\mathcal{M}_f$ and let $P(1)$ be the image in $H^1(K, T_p(\mathcal{M}_f))$ of the Heegner cycle associated with K under the p-adic Abel-Jacobi map. If $P(1)$ is not torsion, then the p-part of the Tate-Šafarevič group of $\mathcal{M}_f$ over K, $Ш_p(\mathcal{M}_f/K)$, is finite.*

[1] Partially supported by an NSF grant

We remark that in [Nek92] there is a stronger condition on p for the theorem to hold which is removed in a remark on the last paragraph of [Nek95].

The purpose of this note is to give the following refinement of the above result:

Theorem 1.2. *There is a finite set of primes $\Psi(f)$, depending only on f, such that for a prime p not in $\Psi(f)$ the following holds: for K as in theorem 1.1, if $P(1)$ is not torsion, then $p^{2\mathcal{I}_p}Ш_p(\mathcal{M}_f/K)=0$, where $\mathcal{I}_p$ is the smallest non-negative integer such that the reduction of $P(1)$ to $H^1(K,T_p(\mathcal{M}_f)/p^{\mathcal{I}_p+1})$ is not 0. In particular, if $\mathcal{I}_p=0$, then $Ш_p(\mathcal{M}_f/K)$ is trivial*

Remark 1.3.
1. The Tate-Šafarevič group discussed here is not exactly the same as the one that appears in [Nek92]. The main difference is in the local conditions at the primes of bad reduction. Nekovář makes no conditions at these primes, which is why Ш comes out too big. The local condition that we use is the one defined by Bloch and Kato. The analysis of this local condition is one of the main ingredient of this work.
2. The finite set $\Psi(f)$ contains the primes dividing $2N$ and primes with an exceptional image of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ in $\mathrm{Aut}(T_p(\mathcal{M}_f))$ (see definition 6.1).

It is our hope that the methods used here allow a complete analysis of the structure of $Ш_p(\mathcal{M}_f/K)$ in terms of various Kolyvagin classes following [Kol91, McC91]. Notice however that some difficulties are already visible in the fact that the power of p annihilating Ш is $2\mathcal{I}_p$ whereas in the elliptic curves case one gets annihilation by $p^{\mathcal{I}_p}$. This difficulty is caused by the more complicated structure of the image of the Galois representation associated to $\mathcal{M}_f$ (see remark 6.5).

A natural problem raised by theorem 1.2 is to bound the numbers $\mathcal{I}_p$. In particular, one would hope that $\mathcal{I}_p=0$ for all but a finite number of p's. This would show the finiteness of $Ш(\mathcal{M}_f/K)$ except for possible infinite contribution at primes dividing $2N$. It is useful to compare the situation to the case where the weight of f is 2, where the triviality of $Ш_p(\mathcal{M}_f/K)$ for almost all p has been previously established in [KL90]. In that case, the class $P(1)$ correspond to a point on the Jacobian of a modular curve, and $\mathcal{I}_p=0$ for almost all p whenever $P(1)$ is of infinite order. This last result uses essentially the injectivity of the Abel-Jacobi map (up to torsion) and the Mordell-Weil theorem, neither of which is known for greater than 1 codimension cycles. One possible way of getting some control over the indices $\mathcal{I}_p$ could be to use the results of Nekovář on the p-adic heights of Heegner cycles: According to [Nek95, corollary to theorem A] one has the equality $h(P(1),P(1))=\Omega_{f\otimes K,p}L'_p(f\otimes K,r)$ where $h(\ ,\ )$ is the p-adic height pairing defined by Nekovář and Perrin-Riou, $L_p(f\otimes K)$ is a p-adic L-function of f over K defined by Nekovář and $\Omega_{f\otimes K,p}$ is some p-adic period. The p-adic height of elements of $H^1_f(K,T)$ has a bounded denominator (it is integral for universal norms from a $\mathbb{Z}_p$ extension) and so the estimation of $\mathcal{I}_p$ is reduced to giving estimates on the p-divisibility of $L'_p(f\otimes K,r)$.

Another problem is to handle primes dividing $2N$. The difficulty here is that we do not understand yet the image of the Abel-Jacobi map with $\mathbb{Q}_p$ coefficients for varieties over an extension of $\mathbb{Q}_p$ and with bad reduction. Recently there has been some progress on that problem [Lan96] but the results do not yet cover the cases we need.

Here is a short description of the contents. After a few preliminary remarks and definitions in section 2 we will recall in section 3 some of the main points of [Nek92].

For brevity this will be far from a full account. We merely attempt to indicate the main changes that need to be made and explain where the local conditions at the bad primes come into play. These conditions are then discussed in sections 4 and 5. We then give the proof of the main theorem in section 6. It would have been nice to skip this section or make it shorter and refer instead to the corresponding sections in [Nek92]. However, it turns out that to get the result we want under weaker conditions than the ones stated there (see the remark in loc. cit. page 121), the proof has to be modified somewhat. I have therefore chosen to give the full details of the proof. In the appendix we give a proof of a Hochschild-Serre spectral sequence for continuous group cohomology which is used in section 5.

As the reader will notice, this work is closely related to [Nek92]. Familiarity with that paper is helpful for reading this one but not necessary, as one may choose to trust the results quoted from there.

I would like to thank Wayne Raskind, Don Blasius, Haruzo Hida, Dinakar Ramakrishnan and Jan Nekovář for helpful discussions and remarks. I would also like to thank Farshid Hajir for encouraging me to write down my ideas on this subject. Finally, I would like to thank the referee for some useful corrections and remarks.

2 PRELIMINARIES

For this work, a motive is effectively equivalent to its set of realizations. We only need the p-adic realizations for the different p's and a brief mention of the Betti realization. Thus, a motive $\mathcal{M}$ has a Betti realization which is a $\mathbb{Q}$-vector space $V_{\mathbb{Q}}$ and p-adic realizations which are continuous representations of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ on $V_p = V_{\mathbb{Q}} \otimes \mathbb{Q}_p$ for the different p's. By choosing a suitable $\mathbb{Z}$-lattice $T_{\mathbb{Z}}$ in $V_{\mathbb{Q}}$ we have in each V_p an invariant $\mathbb{Z}_p$-lattice $T_p = T_{\mathbb{Z}} \otimes \mathbb{Z}_p$. The p-part of the Tate-Šafarevič group of $\mathcal{M}$ depends on the choice of T_p but statements about the p-part for all but a finite number of p are clearly independent of the choice of $T_{\mathbb{Z}}$. In the cases we will be considering there is a standard choice (a Tate twist of a piece of the étale cohomology of a suitable Kuga-Sato variety, see [Nek92, §3]) and the theorem will be proved for this choice. To be more precise:

$$T_p \otimes \mathbb{Q}_p \cong \rho_{f,p} \otimes \mathbb{Q}_p(r), \tag{2.1}$$

where $\rho_{f,p}$ is the standard p-adic representation associated to f.

To define the p-part of Ш, we start with the free $\mathbb{Z}_p$-module of finite rank, $T = T_p(\mathcal{M})$, on which $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ acts continuously. Let $V = T \otimes \mathbb{Q}_p$ and $A = V/T$, so that there is a short exact sequence:

$$0 \to T \xrightarrow{i} V \xrightarrow{\mathrm{pr}} A \to 0.$$

Let ℓ be a prime, possibly ∞. Let F be a finite extension of $\mathbb{Q}_\ell$ and let $\bar{F}$ be an algebraic closure of F. In [BK90, (3.7.1)] Bloch and Kato define the finite part H^1_f of the first Galois cohomology of F with values in V, T or A as follows:

$$\begin{aligned}
&H^1_f(F,V) := \operatorname{Ker} H^1(F,V) \xrightarrow{\text{res}} H^1(F^{ur},V) \text{ when } \ell \neq p;\\
&H^1_f(F,V) := \operatorname{Ker} H^1(F,V) \to H^1(F, V\otimes B_{cris}) \text{ when } \ell = p;\\
&H^1_f(F,T) := i^{-1}H^1_f(F,V);\\
&H^1_f(F,A) := \operatorname{Im} H^1_f(F,V) \hookrightarrow H^1(F,V) \xrightarrow{\text{pr}} H^1(F,A),
\end{aligned}$$

where F^{ur} is the maximal unramified extension of F. The ring B_{cris} is defined by Fontaine. We will not need to use the definition directly in the case $\ell = p$.

Let now K be a number field. When B is a $\operatorname{Gal}(\bar{\mathbb{Q}}/K)$-module we have restriction maps for each place v of K: $H^1(K,B) \to H^1(K_v,B)$. When $x \in H^1(K,B)$ we will denote its restriction to $H^1(K_v,B)$ by x_v. The p-part of the Selmer group of $\mathcal{M}$ over K is now defined as

$$\operatorname{Sel}_p(\mathcal{M}/K) := \operatorname{Ker} H^1(K,A) \longrightarrow \prod_v H^1(K_v,A)/H^1_f(K_v,A),$$

where the product is over all places v of K. We also define

$$H^1_f(K,V) := \operatorname{Ker} H^1(K,V) \longrightarrow \prod_v H^1(K_v,V)/H^1_f(K_v,V).$$

The p-part of the Tate-Šafarevič group of $\mathcal{M}$ over K is the quotient of $\operatorname{Sel}_p(\mathcal{M}/K)$ by the image of $H^1_f(K,V)$. Nekovář defines the same group as the quotient of the Selmer group by the image of an appropriate Abel-Jacobi map. It follows easily from his result that in the case of interest here his definition coincides with the one we are using.

Let A_{p^k} be the p^k-torsion subgroup of A and let $\operatorname{red}_{p^k} : T \to A_{p^k}$ be the reduction mod p^k. We will use the same notation for the reduction map $A_{p^n} \to A_{p^k}$ which is given by multiplication by p^{n-k} when $n > k$ and we notice that all reduction maps commute with each other. We will abuse the notation further to denote by red_{p^k} the maps induced by the reduction on Galois cohomology groups.

To simplify the notation slightly, we assume the following:

ASSUMPTION 2.1. There is a Galois invariant bilinear pairing $T \times T \to \mathbb{Z}_p(1)$ such that the induced pairings on $T/p^k \cong A_{p^k}$ are non-degenerate for all k.

This condition is satisfied in the case we are considering by [Nek92, proposition 3.1]. It is mostly made at this point so that we do not have to consider both T and its Kummer dual. We have the following well known results:

PROPOSITION 2.2. *The pairing above induces local Tate pairings, for each place v of K:*

$$\begin{aligned}
H^1(K_v,T) \times H^1(K_v,A) &\to H^1(K_v,\mathbb{Q}_p/\mathbb{Z}_p(1)) \cong \mathbb{Q}_p/\mathbb{Z}_p;\\
H^1(K_v,A_{p^k}) \times H^1(K_v,A_{p^k}) &\to H^1(K_v,\mathbb{Z}/p^k(1)) \cong \mathbb{Z}/p^k,
\end{aligned}$$

which are both perfect and will be denoted by $\langle\ ,\ \rangle_v$ (for the torsion coefficients case see [Mil86, Chap. I, Cor. 2.3]*). The following properties hold:*

1. [BK90, Proposition 3.8] *The pairing* $\langle\ ,\ \rangle_v$ *makes* $H^1_f(K_v,T)$ *and* $H^1_f(K_v,A)$ *exact annihilators of each other (this is true even in the case* $p|v$*).*

2. *If* x *and* y *belong to* $H^1(K,A_{p^k})$ *then*

$$\sum_v \langle x_v, y_v\rangle_v = 0,$$

where the sum is over all places v *of* K *but is in fact a finite sum.*

We remark that it is possible to neglect the infinite places in all the discussions if we assume that $p \neq 2$ or if K is totally imaginary. Both conditions will in fact hold.

DEFINITION 2.3. Let F be a local field. We define $H^1_f(F,A_{p^k})$ to be the preimage in $H^1(F,A_{p^k})$ of $H^1_f(F,A)$. We define $H^1_{f^*}(F,A_{p^k})$ to be the annihilator of $H^1_f(F,A_{p^k})$ in $H^1(F,A_{p^k})$ under local Tate duality. We will call the classes in $H^1_{f^*}(F,A_{p^k})$ the *dual finite classes.* We define the singular part of the cohomology as

$$H^1_{sin}(F,A_{p^k}) = H^1(F,A_{p^k})/H^1_{f^*}(F,A_{p^k})$$

(this definition is due to Mazur). If $x \in H^1(F,A_{p^k})$ we denote by x_{sin} its projection on the singular part. When K is a number field we let

$$\mathrm{Sel}(K,A_{p^k}) := \mathrm{Ker}\, H^1(K,A_{p^k}) \longrightarrow \prod_v H^1(K_v,A_{p^k})/H^1_f(K_v,A_{p^k}).$$

LEMMA 2.4. *The group* $H^1_{f^*}(F,A_{p^k})$ *is the image of* $H^1_f(F,T)$ *under the canonical map* $H^1(F,T) \to H^1(F,A_{p^k})$*. There is a perfect pairing, induced by* $\langle\ ,\ \rangle_v$*:*

$$\langle\ ,\ \rangle_v : H^1_f(F,A_{p^k}) \times H^1_{sin}(F,A_{p^k}) \to \mathbb{Z}/p^k$$

Proof. This is a formal consequence of the preceding definition and proposition 2.2. □

For a $\mathrm{Gal}(\bar F/F)$-module B and $\bar F \supset K \supset F$ we denote $B^{\mathrm{Gal}(\bar F/K)}$ by $B(K)$. If B' is a subset of B we denote by $F(B)$ the fixed field of the subgroup of $\mathrm{Gal}(\bar F/F)$ fixing B'.

3 METHOD OF PROOF

The Kolyvagin method, as applied to $\mathcal{M}_f$ by Nekovář, works as follows: Let f have q-expansion $f = \sum a_n q^n$. Let E be the field generated over $\mathbb{Q}$ by the a_i. It is known that E is a totally real finite extension of $\mathbb{Q}$. Let $\mathcal{O}_E$ be the ring of integers of E. As explained in [Nek92, Proposition 3.1], the invariant lattice $T_p(\mathcal{M}_f)$ can be taken to be a free rank 2 module over $\mathcal{O}_E \otimes \mathbb{Z}_p = \prod \mathcal{O}_{E_\mathfrak{p}}$, where the product is over all primes $\mathfrak{p}$ of E dividing p. To prove the result about Ш it is sufficient to choose one such prime $\mathfrak{p}$ and consider only the direct summand of $T_p(\mathcal{M}_f)$ corresponding to $\mathfrak{p}$. This summand will be denoted $T_{f,\mathfrak{p}}$. For the rest of this section we fix $T = T_{f,\mathfrak{p}}$ and let as usual $V = T \otimes \mathbb{Q}_p$ and $A = V/T$.

As the Tate-Šafarevič group is (obvious with the above definition) p-torsion, we wish to show that its part killed by p^k is killed by the fixed power $p^{2\mathcal{I}_p}$ for each k. We look at the short exact sequence

$$0 \to A_{p^k} \to A \xrightarrow{p^k} A \to 0$$

and the induced sequence on cohomology

$$0 \to A(K)/p^k \to H^1(K, A_{p^k}) \to H^1(K, A)_{p^k} \to 0$$

The conditions we will impose on the prime p imply, as we will see in part 2 of proposition 6.3, that $A(K) = 0$, and hence $H^1(K, A)_{p^k} \cong H^1(K, A_{p^k})$. It follows that the preimage in $H^1(K, A_{p^k})$ of $\mathrm{Sel}_p(T/K)$ is $\mathrm{Sel}(K, A_{p^k})$. Since $P(1) \in H^1_f(K, V)$ it will be enough to show that $\mathrm{Sel}(K, A_{p^k})/(\mathcal{O}_{E_\mathfrak{p}}/p^k)P(1)$ is killed by $p^{2\mathcal{I}_p}$.

Choose once and for all a complex conjugation $\tau \in \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. Let $S(k)$ be the set of primes ℓ satisfying:

- $\ell \nmid NDp$;
- ℓ is inert in K;
- p^k divides a_ℓ and $\ell + 1$;
- $\ell + 1 \pm a_\ell$ are not divisible by p^{k+1}.

Remark 3.1. The first 3 conditions are equivalent to $\mathrm{Frob}(\ell)$ and τ being conjugates in $\mathrm{Gal}(K(A_{p^k})/\mathbb{Q})$. The last condition can be arranged for infinitely many ℓ's (see proposition 6.10).

Let n be a product of distinct primes $\ell \in S(k)$. Nekovář associates with n a cohomology class $y_n \in H^1(K_n, T)$, where K_n is the ring class field of K of conductor n. The classes y_n are defined as the images of certain CM cycles under the Abel-Jacobi map of $\mathcal{M}_f$. When $n = m\ell$ the relation

$$\mathrm{cor}_{K_n, K_m}(y_n) = a_\ell y_m$$

holds, as well as some local congruence condition which we will not discuss here.

Let $G_n := \mathrm{Gal}(K_n/K_1)$. Then $G_n = \prod_{\ell | n} G_\ell$. For each prime $\ell \in S(k)$ we associate the element $D_\ell \in \mathbb{Z}[G_\ell]$ which is given by

$$D_\ell = \sum_{i=1}^{\ell} i\sigma^i, \qquad G_\ell = \langle \sigma \rangle,$$

and let $D_n = \prod_{\ell | n} D_\ell \in \mathbb{Z}[G_n]$. One now notices, following Kolyvagin, that $D_n(\mathrm{red}_{p^k} y_n) \in H^1(K_n, A_{p^k})$ is G_n-invariant. By [Nek92, Proposition 6.3]

$$p^M A_{p^k}(K_n) = 0, \tag{3.1}$$

with some constant M independent of n and k. An application of the inflation restriction sequence shows that there is a canonically defined class $z_n \in H^1(K_1, A_{p^{k-2M}})$ such that

$$\mathrm{res}_{K_1, K_n} z_n = D_n(\mathrm{red}_{p^{k-2M}} y_n).$$

Indeed, one has the commutative diagram with exact inflation restriction rows:

$$\begin{array}{ccccc}
H^1(K_1, A_{p^k}) & \xrightarrow{\mathrm{res}_{K_1,K_n}} & H^1(K_n, A_{p^k})^{G_n} & \longrightarrow & H^2(G_n, A_{p^k}(K_n)) \\
{\scriptstyle \mathrm{red}_{p^{k-M}}}\downarrow & & {\scriptstyle \mathrm{red}_{p^{k-M}}}\downarrow & & {\scriptstyle \mathrm{red}_{p^{k-M}}}\downarrow \\
H^1(K_1, A_{p^{k-M}}) & \xrightarrow{\mathrm{res}_{K_1,K_n}} & H^1(K_n, A_{p^{k-M}})^{G_n} & \longrightarrow & H^2(G_n, A_{p^{k-M}}(K_n))
\end{array}$$

and the rightmost vertical map is 0 by (3.1) because the reduction map kills p^M torsion. It follows that

$$\mathrm{red}_{p^{k-M}}\, y_n \in \mathrm{Im}\left(\mathrm{res}_{K_1,K_n} : H^1(K_1, A_{p^{k-M}}) \to H^1(K_n, A_{p^{k-M}})\right).$$

We get the canonical class z_n by further reduction as in [Nek92, §7]. Finally, define

$$P(n) := \mathrm{cor}_{K_1,K}\, z_n.$$

Note the important difference between Nekovář's definition of the same classes and ours: in Nekovář's definition $\mathrm{res}_{K_1,K_n} z_n = p^M D_n(\mathrm{red}_{p^{k-M}}\, y_n)$. To simplify the notation, we may notice that the definition is entirely independent of the value of M. To define classes in the cohomology of A_{p^r} we need to start with n whose prime divisors satisfy certain congruences depending on r and M and we may freely assume that we have chosen the n correctly whatever the congruences are. It will be convenient to make the change of variable $k = k - 2M$ here. Note that $P(1)$ can be considered mod p^k for any k and its definition is independent of M.

Proposition 3.2. *The classes $P(n)$ enjoy certain fundamental properties:*

1. *$P(n)$ belongs to the $(-1)^{par(n)}\varepsilon_L$-eigenspace of the complex conjugation τ acting on $H^1(K, A_{p^k})$, where $par(n)$ is the parity of the number of prime factors in n and ε_L is the negative of the sign of the functional equation of $L(f,s)$.*

2. *For a place v of K such that $v \nmid Nn$, $P(n) \in H^1_{f^*}(K_v, A_{p^k})$.*

3. *If $n = m\cdot\ell$ and λ is the unique prime of K above ℓ, then there is an isomorphism between $H^1_f(K_\lambda, A_{p^k})$ and $H^1_{sin}(K_\lambda, A_{p^k})$ which takes $P(m)_\lambda$ to $P(n)_{\lambda,sin}$. In particular, if $P(m)_\lambda \neq 0$, then $P(n)_{\lambda,sin} \neq 0$.*

Proof. This is [Nek92, Proposition 10.2] with a couple of modifications. First of all we remark that there is a miss-print in [Nek92] and the eigenvalue of τ on $P(n)$ is indeed $(-1)^{par(n)}\varepsilon_L$ as can be seen from the proof. To get the second statement when $v \nmid p$ we note that if such a v is a prime of good reduction one has $H^1_{f^*}(K_v, A_{p^k}) = H^1_f(K_v, A_{p^k}) = H^1_{ur}(K_v, A_{p^k})$ (see lemma 4.4) and that the auxiliary power of p that appear in [Nek92] is not needed here because of the change in the definition of $P(n)$ alluded to above. The case $v|p$ follows from [Nek92, Lemma 11.1]. Here, two remarks are in place: First of all, Nekovář uses the comparison theorem of Faltings for open varieties [Fal89]. As is well known, this result is not universally accepted. However, in the last 2 years Nekovář himself [Nek96] and Nizioł [Niz97, Theorem 3.2] have supplied alternative proofs that the image of the Abel-Jacobi map lies inside H_f in the case of good reduction. The second remark is that this is all we need because our assumption $p \nmid 2N$ imply that $v|p$ is a place of good reduction. $\square$

One of the main points of this work is to analyze the dual finite conditions at primes of bad reduction and to show that by further reduction (i.e. by possibly increasing M) one may assume that the classes $P(n)$ are dual finite at these primes (see corollary 5.2).

4 Finite and dual finite conditions at ℓ

Let F be a finite extension of $\mathbb{Q}_\ell$ ($\ell \neq p$) and let T be a free $\mathbb{Z}_p$-module of finite rank with a continuous action of $G = \mathrm{Gal}(\bar{F}/F)$. Again let $V = T \otimes \mathbb{Q}_p$ and $A = V/T$. Let $I = \mathrm{Gal}(\bar{F}/F^{ur})$ be the inertial group. We assume the following condition is satisfied (as is in the case at hand, see [Nek92, proposition 3.1]):

Condition 4.1. There is a Galois invariant, non-degenerate bilinear pairing $V \times V \to \mathbb{Q}_p(1)$ and $V_I(-1)$ has no nontrivial fixed vector with respect to any power of Frobenius (true if V_I has no part of weight -2).

Proposition 4.2. *Under the above condition there exists a constant M such that for any finite unramified extension L/F we have*

1. $p^M H^1(L^{ur}, T)^{\mathrm{Gal}(L^{ur}/L)} = 0$;
2. $H^1_f(L, V) = H^1(L, V)$;
3. $V(L) = 0$.

Proof. The second statement immediately follows from the first. For the first statement we begin by noticing that I is independent of L. By making a finite ramified extension we may assume that the action of I factors through the p-primary part of its tame quotient. It then follows that $H^1(I, T) \cong T_I(-1)$ as $\mathrm{Gal}(L^{ur}/L)$-modules. The condition now implies that $T_I(-1)$ is a direct sum of a torsion group and a $\mathbb{Z}_p$-free module on which Frobenius has no invariants. Finally, the third statement follows since by duality one gets that 1 is not an eigenvalue of any power of Frobenius on V^I. □

Remark 4.3. If T is the Tate module of an elliptic curve with split semi-stable reduction, then the constant M is essentially the p-adic valuation of the number of components of the special fiber of E.

It follows from part 2 of proposition 4.2 that for any finite unramified extension L/F we have $H^1_f(L, T) = H^1(L, T)$, and therefore by lemma 2.4 we get

$$H^1_{f^*}(L, A_{p^k}) = \mathrm{Im}\, H^1(L, T) \xrightarrow{\mathrm{red}} H^1(L, A_{p^k}).$$

Lemma 4.4. *If the G-module T is unramified, then for any L as above*

$$H^1_{f^*}(L, A_{p^k}) = H^1_f(L, A_{p^k}) = H^1_{ur}(L, A_{p^k}) := \mathrm{Ker}\, H^1(L, A_{p^k}) \to H^1(L^{ur}, A_{p^k}).$$

Proof. It is enough to show the second equality as the condition of being unramified is self dual. It is clear that any class in $H^1_f(L, A_{p^k})$ is unramified. Conversely, a class in $H^1_{ur}(L, A_{p^k})$ is inflated from $H^1(L^{ur}/L, A_{p^k})$. Since $\mathrm{Gal}(L^{ur}/L) \cong \hat{\mathbb{Z}}$, H^1 is just coinvariants. It follows that the reduction map $H^1(L^{ur}/L, T) \to H^1(L^{ur}/L, A_{p^k})$ is surjective. □

5 The local condition under restriction

Keeping the assumption of the previous section, suppose now that L/F is a finite unramified extension with Galois group Δ. The short exact sequence $0 \to T \xrightarrow{p^k} T \xrightarrow{\mathrm{red}_{p^k}} A_{p^k} \to 0$ gives rise to the following commutative diagram with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & H^1(F,T)/p^k \arrow[r, "\mathrm{red}_{p^k}"] \arrow[d, "\mathrm{res}_{F,L}"'] & H^1(F,A_{p^k}) \arrow[r] \arrow[d, "\mathrm{res}_{F,L}"'] & H^2(F,T)_{p^k} \arrow[r] \arrow[d, "\mathrm{res}_{F,L}"'] & 0 \\
0 \arrow[r] & \left(H^1(L,T)/p^k\right)^\Delta \arrow[r, "\mathrm{red}_{p^k}"] & H^1(L,A_{p^k})^\Delta \arrow[r] & H^2(L,T)^\Delta_{p^k} &
\end{tikzcd} \tag{5.1}$$

Given $x \in H^1(F,A_{p^k})$ such that $\mathrm{res}_{F,L}\, x$ is in $H^1_{f^*}(L,A_{p^k})$, we would like to know how far is x from being in $H^1_{f^*}(F,A_{p^k})$. In view of (5.1) the obstruction is given by

$$\operatorname{Ker} H^2(F,T)_{p^k} \xrightarrow{\mathrm{res}_{F,L}} H^2(L,T)^\Delta_{p^k}. \tag{5.2}$$

Proposition 5.1. *The kernel* (5.2) *is annihilated by a constant p^M independent of k and L.*

Proof. Since Δ is finite, there is a Hochschild-Serre spectral sequence

$$E_2^{i,j} = H^i(\Delta, H^j(L,T)) \Rightarrow H^{i+j}(F,T).$$

Note that the cohomology here is the continuous cohomology. The Hochschild-Serre spectral sequence does not exist in general for continuous cohomology. A proof that it does exits in our case is found in the appendix. For $i+j=2$ the spectral sequence converges to a filtration $F^0 \supset F^1 \supset F^2 \supset 0$ on $H^2(F,T)$ with

$$\begin{aligned}
F^1 &= \operatorname{Ker} H^2(F,T) \xrightarrow{\mathrm{res}_{F,L}} H^2(L,T)^\Delta;\\
F^1/F^2 &\cong E_\infty^{1,1} = E_3^{1,1} = \operatorname{Ker}\left[H^1(\Delta, H^1(L,T)) \to H^3(\Delta, T(L))\right]\\
&= H^1(\Delta, H^1(L,T));\\
F^2 &\cong E_\infty^{2,0} \subset E_2^{2,0} = H^2(\Delta, T(L)) = 0,
\end{aligned}$$

since $T(L)=0$ by part 3 of proposition 4.2. Therefore,

$$\operatorname{Ker}\left(H^2(F,T)_{p^k} \xrightarrow{\mathrm{res}_{F,L}} H^2(L,T)^\Delta_{p^k}\right) \cong H^1(\Delta, H^1(L,T))_{p^k}.$$

Applying the inflation restriction sequence to $\mathrm{Gal}(L^{ur}/L) \triangleleft \mathrm{Gal}(\bar{L}/L)$ and T we find

$$0 \to H^1(L^{ur}/L, T(L^{ur})) \to H^1(L,T) \to H^1(L^{ur},T)^{\mathrm{Gal}(L^{ur}/L)} \to 0.$$

The right exactness is a consequence of the fact that $\mathrm{Gal}(L^{ur}/L) \cong \hat{\mathbb{Z}}$ has cohomological dimension 1. Applying the Hochschild-Serre spectral sequence to $\mathrm{Gal}(L^{ur}/L) \triangleleft \mathrm{Gal}(L^{ur}/F)$ and $T(L^{ur})$ we find that $H^1(\Delta, H^1(L^{ur}/L, T(L^{ur})))$ injects into $H^2(L^{ur}/F, T(L^{ur}))$ and is therefore 0 since $\mathrm{Gal}(L^{ur}/F) \cong \hat{\mathbb{Z}}$. Therefore, $H^1(\Delta, H^1(L,T)) \hookrightarrow H^1(\Delta, H^1(L^{ur},T)^{\mathrm{Gal}(L^{ur}/L)})$ and the result follows from proposition 4.2 $\square$

COROLLARY 5.2. *Let p^M be the constant given by proposition 5.1. Then, if $x \in H^1(F, A_{p^{k+M}})$ and $\mathrm{res}_{F,L}\, x \in H^1_{f^*}(L, A_{p^{k+M}})$, then $\mathrm{red}_{p^k}\, x \in H^1_{f^*}(F, A_{p^k})$.*

Proof. The commuting diagram with exact rows

$$\begin{tikzcd}
0 \arrow[r] & T \arrow[r, "p^{k+M}"] \arrow[d, "p^M"'] & T \arrow[r, "\mathrm{red}_{p^{k+M}}"] \arrow[d, "="'] & A_{p^{k+M}} \arrow[r] \arrow[d, "\mathrm{red}_{p^k}"'] & 0 \\
0 \arrow[r] & T \arrow[r, "p^k"] & T \arrow[r, "\mathrm{red}_{p^k}"] & A_{p^k} \arrow[r] & 0
\end{tikzcd}$$

gives rise to

$$\begin{tikzcd}
H^1(F,T) \arrow[r, "\mathrm{red}_{p^{k+M}}"] \arrow[d, "="'] & H^1(F, A_{p^{k+M}}) \arrow[r] \arrow[d, "\mathrm{red}_{p^k}"'] & H^2(F,T)_{p^{k+M}} \arrow[d, "p^M"'] \\
H^1(F,T) \arrow[r, "\mathrm{red}_{p^k}"] & H^1(F, A_{p^k}) \arrow[r] & H^2(F,T)_{p^k}
\end{tikzcd}$$

The corollary now follows by a diagram chase on this last diagram as well as on (5.1) with k replaced by $k + M$. □

6 PROOF OF THEOREM 1.2

In this section we give the proof of the main theorem using a variant of the Kolyvagin argument following mostly [Gro91]. By proposition 3.2 and corollary 5.2 we may assume that the class $P(n)$ is dual finite at all primes which do not divide n. Recall that this involves fixing some large integer M, constructing the classes modulo p^{k+M} and then reducing them mod p^k.

We will concentrate on the case where f has no CM. The CM case can be handled similarly (see the remark in [Nek92] page 121). Recall that E is the field generated by the Fourier coefficients of the form f. We first exclude primes p which are ramified in E. If p is not excluded, let $\mathfrak{p}$ be a prime of E above p and recall that we are considering $T = T_{f,\mathfrak{p}}$ which is a rank 2 free $\mathcal{O}_{E_\mathfrak{p}}$-module with an action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. Let again $\rho_{f,p}$ be the p-adic representation associated with f. Consider the $\mathfrak{p}$ component of $\rho_{f,p}$ which is a representation of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ on a 2-dimensional $E_\mathfrak{p}$ vector space $V_{\rho_f,\mathfrak{p}}$. According to a result of Ribet [Rib85, theorem 3.1] if p is outside a finite set of primes then there is a subfield E' of $E_\mathfrak{p}$ such that in an appropriate basis the image of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ in $\mathrm{Aut}(V_{\rho_f,\mathfrak{p}}) \cong \mathrm{GL}_2(E_\mathfrak{p})$ contains

$$\{g \in \mathrm{GL}_2(\mathcal{O}_{E'}),\ \ \det g \in ((\mathbb{Z}_p^\times)^{2r-1})\}$$

(in fact, the result of Ribet is stronger and treats the image of Galois in all the completions of E above p simultaneously), and therefore contains in particular

$$\{g \in \mathrm{GL}_2(\mathbb{Z}_p),\ \ \det g \in ((\mathbb{Z}_p^\times)^{2r-1})\}. \tag{6.1}$$

We exclude all other primes and the prime 2. This concludes our exclusions which we may sum up in:

DEFINITION 6.1. The set $\Psi(f)$ of excluded primes for theorem 1.2 is the set containing the primes dividing $2N$, primes that ramify in $E = \mathbb{Q}(a_i)$ and primes where the image of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ in $\mathrm{Aut}(V_{\rho_f,\mathfrak{p}})$ does not contains (6.1) (in some basis).

We consider non excluded primes from now onward.

LEMMA 6.2. *Let $\tilde{G}_p$ be the image of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ in $\mathrm{Aut}(T) \cong \mathrm{GL}_2(\mathcal{O}_{E_\mathfrak{p}})$ (p not excluded). Then, $\tilde{G}_p$ contains a subgroup conjugate to $\mathrm{GL}_2(\mathbb{Z}_p)$.*

Proof. By (2.1), $T \otimes E_\mathfrak{p}$ is just the r-th Tate twist of $V_{\rho_f,\mathfrak{p}}$. From that and Ribet's theorem it follows easily that after fixing an appropriate basis for T every matrix $A \in \mathrm{GL}_2(\mathcal{O}_{E_\mathfrak{p}})$ has a scalar multiple in $\tilde{G}_p$. Since $\mathrm{SL}_2(\mathcal{O}_{E_\mathfrak{p}})$ is the commutator subgroup of $\mathrm{GL}_2(\mathcal{O}_{E_\mathfrak{p}})$, it follows that $\mathrm{SL}_2(\mathcal{O}_{E_\mathfrak{p}}) \subset \tilde{G}_p$. The lemma follows because for almost all ℓ, $\mathrm{Frob}(\ell)$ has determinant ℓ^{-1} and because $\tilde{G}_p$ is closed. □

Let $\mathbb{F} = \mathcal{O}_{E_\mathfrak{p}}/p^k$. Let $G_{p^k} \cong \mathrm{Gal}(\mathbb{Q}(A_{p^k})/\mathbb{Q})$ be the image of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ in $\mathrm{Aut}(A_{p^k}) \cong \mathrm{GL}_2(\mathbb{F})$. Then, G_{p^k} contains a group G'_{p^k} conjugate to $\mathrm{SL}_2(\mathbb{Z}/p^k)$.

PROPOSITION 6.3. *Let $L = K(A_{p^k})$.*

1. *When $k = 1$, A_p is an irreducible $\mathbb{F}[\mathrm{Gal}(L/K)]$-module.*
2. *$H^i(\mathrm{Gal}(L/K), A_{p^k}) = 0$ for all $i \geq 0$.*
3. *There is a natural pairing $[\ ,\ ] : H^1(K, A_{p^k}) \times \mathrm{Gal}(\bar{\mathbb{Q}}/L) \to A_{p^k}$ inducing an isomorphism of $\mathbb{F}$-modules $H^1(K, A_{p^k}) \cong \mathrm{Hom}_{\mathrm{Gal}(L/K)}(\mathrm{Gal}(\bar{\mathbb{Q}}/L), A_{p^k})$.*
4. *The $\mathbb{F}$-module A_{p^k} is the direct sum of its ± 1 eigenspaces with respect to the generator τ of $\mathrm{Gal}(K/\mathbb{Q})$, each free of rank 1.*

Proof. Since $\mathrm{SL}_2(\mathbb{F}_p)$ has no nontrivial $\mathbb{Z}/2$ quotients when $p > 2$ and $\mathrm{Gal}(L/K)$ is of index at most 2 in G_p, it follows that $\mathrm{Gal}(L/K)$ contains G'_p and therefore that A_p is an irreducible $\mathbb{F}[\mathrm{Gal}(L/K)]$-module. It also follows that $\mathrm{Gal}(L/K)$, considered as embedded in $\mathrm{Aut}(A_{p^k})$, contains the central Subgroup of order 2 generated by -1. Since $p \neq 2$, $H^i(\pm 1, A_{p^k}) = 0$ for all $i \geq 0$ and the second assertion follows from the Hochschild-Serre spectral sequence $H^i(\mathrm{Gal}(L/K)/\pm 1, H^j(\pm 1, A_{p^k})) \Rightarrow H^{i+j}(\mathrm{Gal}(L/K), A_{p^k})$. An inflation restriction sequence now implies that

$$H^1(K, A_{p^k}) \cong H^1(L, A_{p^k})^{\mathrm{Gal}(L/K)} \cong \mathrm{Hom}_{\mathrm{Gal}(L/K)}(\mathrm{Gal}(\bar{\mathbb{Q}}/L), A_{p^k})$$

hence the third assertion. Finally, part 4 follows because the determinant of τ on T is -1. □

Let S be a finitely generated $\mathbb{F}$-submodule of $H^1(K, A_{p^k})$. We consider the elements of S as elements of $\mathrm{Hom}_{\mathrm{Gal}(L/K)}(\mathrm{Gal}(\bar{\mathbb{Q}}/L), A_{p^k})$ and let L_S be the field fixed by the common kernel of these elements. The following lemma is immediate:

LEMMA 6.4. *The pairing $[\ ,\ ]$ induces a pairing*

$$[\ ,\ ]_S : S \times \mathrm{Gal}(L_S/L) \to A_{p^k},$$

which in turn induces an injection

$$\mathrm{Gal}(L_S/L) \hookrightarrow \mathrm{Hom}_{\mathbb{F}}(S, A_{p^k}) \text{ as } \mathrm{Gal}(L/K)\text{-modules.} \tag{6.2}$$

This injection has the property that

$$x \in S \text{ and } [x, \mathrm{Gal}(L_S/L)]_S = 0 \Longrightarrow x = 0.$$

In addition, this pairing induces an injection

$$S \hookrightarrow \mathrm{Hom}_{\mathrm{Gal}(L/K)}(\mathrm{Gal}(L_S/L), A_{p^k}) \text{ as } \mathbb{F}\text{-modules}$$

Remark 6.5. Unlike the situation for elliptic curves [Gro91, proposition 9.3] we can not in general expect the injection (6.2) to be an isomorphism. For instance, if G_{p^k} is contained in $\mathrm{GL}_2(\mathbb{Z}/p^k)$, then there might exist a homomorphism $\phi : \mathrm{Gal}(\bar{\mathbb{Q}}/L) \to A_{p^k}$ whose image is contained in $(\mathbb{Z}/p^k)^2$. If we take S to be the $\mathbb{F}$-span of ϕ, then $\mathrm{Gal}(L_S/L) \cong (\mathbb{Z}/p^k)^2$ and is not in general an $\mathbb{F}$-module whereas $\mathrm{Hom}_{\mathbb{F}}(S, A_{p^k})$ is. The failure of (6.2) to be an isomorphism forces some changes in the final arguments.

Our chosen complex conjugation τ acts on all the groups above. We will denote by $G^{\pm}$ the ± 1-eigenspace of τ acting on an abelian group G.

Lemma 6.6. *Let $C \subset \mathrm{Hom}_{\mathbb{F}}(S, A_{p^k})$ be a $\mathrm{Gal}(L/K)$-submodule with the property that $x \in S$ and $[x, C]_S = 0$ imply $x = 0$. Let $0 \neq s \in S$ and let $a \in \mathrm{Hom}_{\mathbb{F}}(S, A_{p^k})^+$. Let*

$$C' = a + C^+, \ C'' = \{c \in C', \ [s, c]_S \neq 0\}.$$

Then, C' and C'' have the same property as C with respect to eigenvectors of τ in S, that is, if $x \in S^{\pm}$ and $[x, C']_S = 0$ or $[x, C'']_S = 0$, then $x = 0$.

Proof. Suppose first that $[x, C^+]_S = 0$. Then $\mathbb{F} \cdot [x, C]_S$ is an $\mathbb{F}[\mathrm{Gal}(L/K)]$-submodule of A_{p^k} which is contained in the proper submodule $A_{p^k}^{\mp}$. Considering p-torsion and using part 1 of proposition 6.3 one finds that $\mathbb{F} \cdot [x, C]_S$ is trivial. It follows in particular that $[s, C^+]_S$ is non trivial and since $p \geq 3$ it contains at least 3 elements. From that it follows that for any $c \in C^+$ one may always find $c_1, \ c_2 \in C^+$ such that $c = (a + c_1) - (a + c_2)$ and $[s, a + c_i]_S \neq 0$ for $i = 1, 2$. The lemma follows easily. $\square$

Lemma 6.7. *Let ℓ be a prime in $S(M+k)$. Then, ℓ is inert in K. Let λ be the unique prime of K above ℓ. Then, for any choice of $\mathrm{Frob}(\lambda)$ in a decomposition group of λ, $\mathrm{Frob}(\lambda)$ acts trivially on A_{p^k} and therefore λ splits completely in L.*

Proof. Both assertions follow from remark 3.1. In $\mathrm{Gal}(K/\mathbb{Q})$, $\mathrm{Frob}(\ell) = \tau$ hence ℓ is inert in K. It now follows that $\mathrm{Frob}(\lambda)$ is conjugate to τ^2 and is therefore the identity on A_{p^k}. $\square$

Let ℓ and λ be as in the previous lemma, let λ' be a prime of L_S above λ and let $\mathrm{Frob}(\lambda') \in \mathrm{Gal}(L_S/L)$ be the associated Frobenius substitution. It is easy to see that the formula

$$\phi_{\lambda'}(x) := [x, \mathrm{Frob}(\lambda')]_S$$

defines an element of $\mathrm{Hom}_{\mathbb{F}}(S, A_{p^k})$ which depends only on ℓ up to conjugation on A_{p^k} by some element of $\mathrm{Gal}(L/K)$. Using lemma 6.7 one has:

LEMMA 6.8. *There is a* $\mathrm{Gal}(K/\mathbb{Q})$*-equivariant isomorphism*

$$H^1_f(K_\lambda, A_{p^k}) \cong H^1(K_\lambda^{ur}/K_\lambda, A_{p^k}) \cong A_{p^k}, \tag{6.3}$$

where the last step is evaluation at the Frobenius. If $x \in H^1(K, A_{p^k})$ *and* $x_\lambda \in H^1_f(K_\lambda, A_{p^k})$ *then, up to conjugation as before, the image of* x_λ *under this isomorphism is* $\phi_{\lambda'}(x)$.

LEMMA 6.9. *Let* λ *be as above.*

1. *The pairing* $\langle\ ,\ \rangle_\lambda$ *defined in lemma* 2.4 *induces nondegenerate pairings:*

$$\langle\ ,\ \rangle_\lambda^\pm : H^1_f(K_\lambda, A_{p^k})^\pm \times H^1_{sin}(K_\lambda, A_{p^k})^\pm \to \mathbb{Z}/p^k.$$

2. *Both* $H^1_f(K_\lambda, A_{p^k})$ *and* $H^1_{sin}(K_\lambda, A_{p^k})$ *are direct sums of their* ± 1 *eigenspaces with respect to* τ. *All eigenspaces are free of rank* 1 *over* $\mathbb{F}$.

Proof. The first assertion follows since $\langle\ ,\ \rangle_\lambda$ is $\mathrm{Gal}(L/K)$ equivariant. The second assertion follows for $H^1_f(K_\lambda, A_{p^k})$ by lemma 6.8 and part 4 of proposition 6.3 and the same now follows for $H^1_{sin}(K_\lambda, A_{p^k})^\pm$ by the first assertion. □

PROPOSITION 6.10. *Let* $x,\ y \in S$ *and suppose that* $y \neq 0$. *Then there exists some* $\ell \in S(M+k)$ *such that* $y_\lambda \neq 0$. *If for almost all* $\ell \in S(M+k)$ *with* $y_\lambda \neq 0$ *we have* $x_\lambda = 0$, *then* $x = 0$.

Proof. Let $L_M = K(A_{p^{M+k+1}})$. Let C be the image of $\mathrm{Gal}(\bar{\mathbb{Q}}/L_M)$ in $\mathrm{Gal}(L_S/L)$. We first claim that when considered in $\mathrm{Hom}_{\mathbb{F}}(S, A_{p^k})$, C satisfies the assumption of lemma 6.6. To show that, we first notice that the same argument used to prove that $H^i(\mathrm{Gal}(L/K), A_{p^k}) = 0$ for all $i \geq 0$ in proposition 6.3 shows that $H^i(\mathrm{Gal}(L_M/K), A_{p^k}) = 0$ for all such i. An inflation restriction sequence now shows that

$$\mathrm{Hom}_{\mathrm{Gal}(L/K)}(\mathrm{Gal}(L_M/L), A_{p^k}) = H^1(\mathrm{Gal}(L_M/L), A_{p^k})^{\mathrm{Gal}(L/K)} = 0.$$

This implies that if $x \in S$ satisfies $[x, C]_S = 0$, then in fact $[x, \mathrm{Gal}(L_S/L)]_S = 0$ and the claim follows from lemma 6.4.

By lemma 6.2 the image of $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$ in $\mathrm{Aut}(A_{p^{M+k+1}}) \cong \mathrm{GL}_2(\mathcal{O}_{E_\mathfrak{p}}/p^{M+k+1})$ contains an element of the form $a \cdot I$ such that $a \in 1 + p^{M+k}(\mathbb{Z}/p)^\times$. One checks that this element defines $\rho' \in \mathrm{Gal}(L_M/L_{M-1})$ with the property that if $\mathrm{Frob}(\ell)$ contains $\tau\rho'$, then $\ell \in S(M+k)$.

Now let $L' = L_M \cap L_S$. Then $C = \mathrm{Gal}(L_S/L')$. Consider $\sigma \in C^+$. Since C has odd order we can find $\rho \in C$ such that $\sigma = \rho^\tau \rho$. Let $\rho \cdot \rho' \in \mathrm{Gal}(L_M \cdot L_S/K)$ be the element whose restriction to $\mathrm{Gal}(L_M/K)$ is ρ' and whose restriction to $\mathrm{Gal}(L_S/L')$ is σ. By Čebotarev's density theorem, we may find infinitely many primes ℓ whose Frobenius conjugacy class in $\mathrm{Gal}(L_M \cdot L_S/\mathbb{Q})$ contains $\tau \cdot \rho \cdot \rho'$. Every such ℓ is in $S(M+k)$. In addition, after projecting to $\mathrm{Gal}(L_S/L')$ we find $\mathrm{Frob}(\lambda) = (\tau\rho)^2 = \rho^\tau \cdot \rho = \sigma$. Thus, we are able to generate a full coset of C in $\mathrm{Gal}(L_S/L)$ with these $\mathrm{Frob}(\lambda)$. By lemma 6.8 we are also able to generate all elements σ of this coset for which $[y, \sigma]_S = 0$ with $\{\mathrm{Frob}(\lambda),\ y_\lambda \neq 0\}$. The proposition therefore follows from lemma 6.6. □

Lemma 6.11. *Suppose $x \in \mathrm{Sel}(K, A_{p^k})$ and n is a product of primes in $S(M+k)$.*

1. $\sum_{\ell | n} \langle x_\lambda, P(n)_{\lambda,sin}\rangle_\lambda = 0.$

2. *If x and $P(n)$ are in the same eigenspace for τ, $p^{k-\mathcal{I}-1}P(n)_{\lambda,sin} \neq 0$ and we have $\langle \mathbb{F}x_\lambda, P(n)_{\lambda,sin}\rangle_\lambda = 0$, then $p^{\mathcal{I}}x_\lambda = 0$.*

Proof. 1. This follows from proposition 2.2, lemma 2.4 and the fact that the classes $P(n)$ are dual finite at primes not dividing n.
2. Consider first the case $k=1$ and $\mathcal{I}=0$. The conditions then imply that $\mathbb{F}x_\lambda$ is a proper subspace of an eigenspace of τ on $H^1_f(K_\lambda, A_p)$ which is 1-dimensional over $\mathbb{F}$ by lemma 6.9 and it follows that $\mathbb{F}x_\lambda = 0$. If k is arbitrary but $\mathcal{I}=0$ then $P(n)_{\lambda,sin}$ has a non trivial image in p-cotorsion hence by the previous case $\mathbb{F}x_\lambda$ has trivial p-torsion but this can only happen if $x_\lambda = 0$. Finally, if $\mathcal{I} \neq 0$ the conditions imply that $P(n)_{\lambda,sin} = p^{\mathcal{I}'}P'$ with $\mathcal{I}' \leq \mathcal{I}$ and P' has a non trivial image in p-cotorsion. Since $\langle \mathbb{F}p^{\mathcal{I}'}x_\lambda, P'\rangle_\lambda = 0$ we get from the previous case $p^{\mathcal{I}'}x_\lambda = 0$. □

The proof of theorem 1.2 may now be completed as follows: Let $\mathcal{I} = \mathcal{I}_p$ and let $\mathcal{J} = \mathcal{I}+1$. We assume that $k > \mathcal{I}$ and we want to prove that $p^{2\mathcal{I}}$ kills $\mathrm{Sel}(K, A_{p^k})/\mathbb{F}P(1)$. Our assumption is that $\mathrm{red}_{p^{\mathcal{J}}}\, P(1) \neq 0$ in $H^1(K, A_{p^{\mathcal{J}}})$. On $H^1(K, A_{p^k})$, multiplication by $p^{k-\mathcal{J}}$ factors as the composition of $\mathrm{red}_{p^{\mathcal{J}}}$ with the map $H^1(K, A_{p^{\mathcal{J}}}) \to H^1(K, A_{p^k})$ induced by the inclusion in the short exact sequence $0 \to A_{p^{\mathcal{J}}} \to A_{p^k} \to A_{p^{k-\mathcal{J}}} \to 0$. Since $A_{p^{k-\mathcal{J}}}(K) = 0$, this induced map is injective and we conclude that $p^{k-\mathcal{J}}P(1) \neq 0$. Let $x \in \mathrm{Sel}(K, A_{p^k})$. Suppose first that x is in the opposite eigenspace to $P(1)$, hence in the same eigenspace as $P(\ell)$ for $\ell \in S(M+k)$ by proposition 3.2. Let S be the $\mathbb{F}$-submodule of $H^1(K, A_{p^k})$ generated by x and $P(1)$. Suppose $\ell \in S(M+k)$ is such that $(p^{k-\mathcal{J}}P(1))_\lambda \neq 0$. Then, by part 3 of proposition 3.2, $p^{k-\mathcal{J}}P(\ell)_{\lambda,sin} \neq 0$ and from that and lemma 6.11 it follows that $p^{\mathcal{I}}x_\lambda = 0$. Proposition 6.10 therefore implies that $p^{\mathcal{I}}x = 0$.

Suppose now that x is in the same eigenspace as $P(1)$ and we claim that $p^{2\mathcal{I}}x$ has to be a multiple of $P(1)$. By proposition 6.10 we may find $\ell \in S(M+k)$ such that $(p^{k-\mathcal{J}}P(1))_\lambda \neq 0$. As before, this implies that $p^{k-\mathcal{J}}P(\ell)_{\lambda,sin} \neq 0$ and hence that $p^{k-\mathcal{J}}P(\ell) \neq 0$. Let S be generated by x, $P(1)$ and $P(\ell)$. Since $p^{k-\mathcal{J}}P(1)_\lambda \neq 0$ and both $P(1)_\lambda$ and x_λ are in the free rank 1 $\mathbb{F}$-module $H^1_f(K_\lambda, A_{p^k})^\pm$, it is easy to see that we may find a combination $x' = \alpha P(1) + p^{\mathcal{I}}x \in S$, with $\alpha \in \mathbb{F}$, such that $x'_\lambda = 0$. Consider now $\ell \neq \ell_1 \in S(M+k)$ such that $p^{k-\mathcal{J}}P(\ell)_{\lambda_1} \neq 0$. Then $p^{k-\mathcal{J}}P(\ell\ell_1)_{\lambda_1,sin} \neq 0$, again by part 3 of proposition 3.2. Let $x'' \in \mathbb{F}x'$. Then

$$\langle x''_\lambda, P(\ell\ell_1)_{\lambda,sin}\rangle_\lambda + \langle x''_{\lambda_1}, P(\ell\ell_1)_{\lambda_1,sin}\rangle_{\lambda_1} = 0.$$

Since $x''_\lambda = 0$ we find $\langle x''_{\lambda_1}, P(\ell\ell_1)_{\lambda_1,sin}\rangle_{\lambda_1} = 0$. Lemma 6.11 implies that $p^{\mathcal{I}}x'_{\lambda_1} = 0$. From proposition 6.10 we get $p^{\mathcal{I}}x' = 0$ and so $p^{2\mathcal{I}}x = -\alpha p^{\mathcal{I}}P(1)$.

A THE HOCHSCHILD-SERRE SPECTRAL SEQUENCE IN CONTINUOUS COHOMOLOGY

Here we prove the following result:

PROPOSITION A.1. *Let G be a profinite group, M a continuous module of G which is the inverse limit of discrete G-modules M_n, $n \in \mathbb{N}$, and H a normal subgroup of G with a finite quotient group $\Delta = G/H$. Then there is a Hochschild-Serre spectral sequence*

$$E_2^{i,j} = H^i(\Delta, H^j(H, M)) \Rightarrow H^{i+j}(G, M), \tag{A.1}$$

where the cohomology of M is the continuous cohomology, i.e., the one computed with respect to continuous cochains as in [Tat76].

Proof. The spectral sequence will be derived from the Grothendieck spectral sequence for the composition of the functors $U : \mathcal{A} \to \mathcal{B}$ and $V : \mathcal{B} \to \mathcal{C}$ defined as follows:

- $\mathcal{A}$ is the category of inverse systems $(M_n)_{n\in\mathbb{N}}$ of discrete G-modules;
- $\mathcal{B}$ is the category of Δ-modules and $\mathcal{C}$ of abelian groups;
- U is the functor which takes an inverse system of G-modules (M_n) to $\varprojlim M_n^H$;
- V is the Δ invariants functor.

In this case, $U \circ V$ is the functor which takes (M_n) to $\varprojlim M_n^G$, because taking invariants commutes with taking limits. The i-th right derived functor of $(M_n) \to \varprojlim M_n^G$ was shown by Jannsen [Jan88a] to be the continuous cohomology $H^i(G, \varprojlim M_n)$ and the same holds with G replaced by H. The only thing left to check is that U takes $\mathcal{A}$ injectives to V acyclics, or even to injectives. For this fact, a proof can be given along the lines of the proof of the usual Hochschild-Serre spectral sequence (see for example [HS76, p.303]). One only needs to give a left adjoint $\bar{U}$ to U which preserves monomorphisms and this is easily done: for a Δ-module N, let $\bar{U}(N)$ be the constant inverse system of N considered as a G-module. Now it is very easy to check that

$$\mathrm{Hom}_{\mathcal{A}}(\bar{U}(N), (M_n)) = \mathrm{Hom}_{\mathcal{B}}(N, \varprojlim M_n^H)$$

and so the proof is complete. □

REFERENCES

[BK90] S. Bloch and K. Kato, *L-functions and Tamagawa numbers of motives*, The Grothendieck Festschrift I (Boston), Prog. in Math., vol. 86, Birkhäuser, 1990, pp. 333–400.

[Fal89] G. Faltings, *Crystalline cohomology and p-adic Galois representations*, Algebraic analysis, geometry and number theory (Baltimore) (J.I Igusa, ed.), Johns Hopkins University Press, 1989, pp. 25–80.

[Gro91] B. Gross, *Kolyvagin's work on modular elliptic curves*, L-functions and arithmetic (Cambridge) (J. Coates and M.J. Taylor, eds.), Lond. Math. Soc. Lect. Note Ser., vol. 153, Cambridge University Press, 1991, pp. 235–256.

[HS76] P. J. Hilton and U. Stammbach, *A course in homological algebra*, GTM, vol. 4, Springer, New York Heidelberg Berlin, 1976.

[Jan88a] U. Jannsen, *Continuous étale cohomology*, Math. Ann. 280 (1988), 207–245.

[Jan88b] U. Jannsen, *Mixed motives and algebraic K-theory*, Lect. Notes in Math., vol. 1400, Springer, Berlin Heidelberg New York, 1988.

[KL90] V. Kolyvagin and D. Logachev, *Finiteness of the Shafarevich-Tate and the group of rational points for some modular abelian varieties*, Leningrad Math. J. 1 (1990), 1229–1253.

[Kol91] V. Kolyvagin, *On the structure of Shafarevich-Tate groups*, Algebraic geometry (Chicago IL. 1989) (Berlin Heidelberg New York), Lect. Notes in Math., vol. 1479, Springer, Berlin Heidelberg New York, 1991, pp. 333–400.

[Lan96] A. Langer, *Local points of motives in semistable reduction*, Preprint, 1996.

[McC91] W. McCallum, *Kolyvagin's work on Shafarevich-Tate groups*, L-functions and arithmetic (Cambridge) (J. Coates and M.J. Taylor, eds.), Lond. Math. Soc. Lect. Note Ser., vol. 153, Cambridge University Press, 1991, pp. 295–316.

[Mil86] J.S. Milne, *Arithmetic duality theorems*, Perspectives in Mathematics, vol. 1, Academic Press, Boston, Mass., 1986.

[Nek92] J. Nekovář, *Kolyvagin's method for Chow groups of Kuga-Sato varieties*, Invent. Math. 107 (1992), no. 1, 99–125.

[Nek95] J. Nekovář, *On the p-adic height of Heegner cycles*, Math. Ann. 302 (1995), 609–686.

[Nek96] J. Nekovář, *Syntomic cohomology and p-adic regulators*, 1996, In preparation.

[Niz97] W. Nizioł, *On the image of p-adic regulators*, Invent. Math. 127 (1997), 375–400.

[Rib85] K. Ribet, *On l-adic representations attached to modular forms II*, Glasgow Math. J. 27 (1985), 185–194.

[Sch90] A. Scholl, *Motives associated to modular forms*, Invent. Math. 100 (1990), 419–430.

[Tat76] J. Tate, *Relations between K_2 and Galois cohomology*, Invent. Math. 36 (1976), 257–274.

Amnon Besser
Department of Mathematics
UCLA
Box 951555
Los Angeles, CA 90095-1555
U.S.A.

Selmer Groups and Torsion Zero Cycles on the Selfproduct of a Semistable Elliptic Curve

Andreas Langer

Received: February 25, 1997

Communicated by Ulf Rehmann

Abstract. In this paper we extend the finiteness result on the p-primary torsion subgroup in the Chow group of zero cycles on the selfproduct of a semistable elliptic curve obtained in joint work with S. Saito to primes p dividing the conductor. On the way we show the finiteness of the Selmer group associated to the symmetric square of the elliptic curve for those primes. The proof uses p-adic techniques, in particular the Fontaine-Jannsen conjecture proven by Kato and Tsuji.

1991 Mathematics Subject Classification: Primary 14H52; Secondary 19E15, 14F30.
Key words and phrases: torsion zero cycles, semistable elliptic curve, multiplicative reduction primes, Selmer group of the symmetric square, Hyodo-Kato cohomology.

Introduction.

In this note we extend the main finiteness result on p-primary torsion zero-cycles on the selfproduct of a semistable elliptic curve in [L-S] to primes $p \geq 3$ where E has (bad) multiplicative reduction, at least under a certain standard assumption. In the course of the proof we will also derive the finiteness of the Selmer group of the symmetric square $\mathrm{Sym}^2 H^1(E)(1)$ for these primes. However, this latter result has already been proven, under the additional condition that the Galois representation

$$\varrho_p : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \longrightarrow \mathrm{Aut}(E_p)$$

is absolutely irreducible (here $E_p = E_p(\overline{\mathbb{Q}})$ is the subgroup of p-torsion elements of E), in a much more general context by Wiles in his main paper ([W] Theorem 3.1) for Selmer groups associated to deformation theories.

To state the Theorems, let E be a semistable elliptic curve over $\mathbb{Q}$ with conductor N and let $X = E \underset{\mathbb{Q}}{\times} E$ be its self-product. Consider the Chow group $CH_0(X)$ of zero-cycles on X modulo rational equivalence and let $CH_0(X)\{p\}$ be — for a fixed prime p — its p-primary torsion subgroup. For a prime p dividing N consider the following hypothesis:

H 1) The Gersten-Conjecture holds for the Quillen-(Milnor)-sheaf $\mathcal{K}_2$ on a regular model $\mathcal{X}$ of X over $\mathbb{Z}_p$.

Then we have

THEOREM A: *Let E be a semistable elliptic curve and $p \geq 3$ a prime such that $p \mid N$, i.e., E has (bad) multiplicative reduction at p. Assume that the condition H 1) is satisfied. Then $CH_0(X)\{p\}$ is a finite group.*

Let $A = H^2(\overline{X}, \mathbb{Q}_p/\mathbb{Z}_p(2))$ be the $\mathbb{Q}_p/\mathbb{Z}_p$-realization of the motive $H^2(X)(2)$ with its $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-action. Then we have

THEOREM B: *Let E be a semistable elliptic curve over $\mathbb{Q}$ and $p \geq 3$ a prime such that $p \mid N$. Then the Selmer group $S(\mathbb{Q}, A)$ is finite.*

REMARKS:

— In [L-S] we showed the finiteness of $CH_0(X)\{p\}$ for primes p such that $p \not| 6$ and E has good reduction at p. We also proved that $CH_0(X)\{p\}$ is zero for almost all p. Therefore Theorem A extends this result to bad primes and provides a further step towards a proof that the full torsion subgroup $CH_0(X)_{tors}$ is finite. In order to find a first example where this is true it remains to consider the 2- and 3-primary torsion in $CH_0(X)$.

— The Selmer group $S(\mathbb{Q}, A)$ coincides with $S(\mathbb{Q}, \mathrm{Sym}^2 H^1(\overline{E}, \mathbb{Q}_p/\mathbb{Z}_p(1)))$ that was studied by [Fl], because $S(\mathbb{Q}, \mathbb{Q}_p/\mathbb{Z}_p(1))$ is zero. In [Fl] Flach proved the finiteness of $S(\mathbb{Q}, A)$ for primes $p \geq 5$ such that E has good reduction at p and the representation ϱ_p is surjective. We were able to remove the latter hypothesis by using a rank-argument of Bloch-Kato and reproved Flach's finiteness result for primes p such that $p \not| 6N$ (compare [L-S]). In the proof of Theorem B we combine the criterium of Bloch-Kato with Kolyvagin's argument that was used in Flach's paper. Flach's additional condition on the surjectivity of ϱ_p can be avoided by applying a certain lemma, due to J. Nekovář, that bounds the order of $H^1(\mathrm{Gal}(\mathbb{Q}(E_{p^n})/\mathbb{Q}), (\mathrm{Sym}^2 H^1(\overline{E}, \mathbb{Z}/p^n(1)))(-1))$ independently of n.

The paper is organized as follows:
In the first paragraph we reduce the proof of Theorem A to two Lemmas I and II. Lemma I was already proven in ([L-S], Lemma A). Lemma II is similar to ([L-S], Lemma B), but the statement is different. The difference is caused by the particular semistable situation. In the second paragraph we derive Lemma II and Theorem B from a key proposition that bounds the possible corank (at most 1 !) of the cokernel of the map defining the Selmer group. Finally this proposition is proven in the last paragraph. The methods of the proof are similar to those developed in [L-S]. At the

point where the crystalline conjecture was used in the good reduction case, we now use the Fontaine-Jannsen conjecture (proven by Kato/Tsuji for $p \geq 3$) that relates the log-crystalline cohomology to the p-adic étale cohomology. The role of the syntomic cohomology in the context of Schneider's p-adic points conjecture is now replaced by a semistable analog relating log-syntomic cohomology to $H^1_g(\mathbb{Q}_p, H^2(\overline{X}, \mathbb{Q}_p(2)))$ (compare [L]). When we apply this argument we will also need the computation, due to Hyodo and used by Tsuji, on a filtration on the sheaf of p-adic vanishing cycles in terms of modified logarithmic Hodge-Witt sheaves.

This paper was written during a visit at the University of Cambridge. I want to thank J. Coates and J. Nekovář for their invitation and J. Nekovář for many discussions and the permission to include his proof of Lemma (2.5) in this paper. Finally I thank S. Saito for encouraging me to look at the remaining semistable reduction case of our main finiteness result in [L-S] and I consider this work as having been done very much in the spirit of our joint paper and a continuation of it.

§1

We first fix some notations.

For an Abelian group M let M_{div} be the maximal divisible subgroup of M and $M\{p\}$ its p-primary torsion subgroup. For a scheme Z over a field k let $\overline{Z} = Z \underset{k}{\times} \overline{k}$ where k is an algebraic closure of k. Denote by $G_k = \mathrm{Gal}(\overline{k}/k)$ the absolute Galois group of k. We will consider the Zariski sheaf $\mathcal{K}_2$ associated to the presheaf $U \to \mathcal{K}_2(U)$ of Quillen (-Milnor) K-groups on Z and let $H^j_{Zar}(Z, \mathcal{K}_2)$ be its Zariski cohomology. Let E be a semistable elliptic curve over $\mathbb{Q}$ with conductor N, $\phi : X_0(N) \to E$ a modular parametrization of E, $X = E \underset{\mathbb{Q}}{\times} E$. Let T, A, V be the following $G = G_{\mathbb{Q}}$-modules:

$$T = H^2(\overline{X}, \mathbb{Z}_p(2)) \quad , \quad A = H^2(\overline{X}, \mathbb{Q}_p/\mathbb{Z}_p(2)) \quad , \quad V = H^2(\overline{X}, \mathbb{Q}_p(2)) \quad .$$

Note that as Abelian groups $T \cong \mathbb{Z}_p^6$, $A \cong \mathbb{Q}_p/\mathbb{Z}_p^6$, because the integral cohomology of an Abelian variety is torsion-free and the second Betti number of X b_2 is 6.

Let K be the function field of X. For a prime p let

$$NH^3(X, \mathbb{Q}_p/\mathbb{Z}_p(2)) := \ker(H^3(X, \mathbb{Q}_p/\mathbb{Z}_p(2)) \to H^3(K, \mathbb{Q}_p/\mathbb{Z}_p(2)))$$

and

$$K_N H^3(X, \mathbb{Q}_p/\mathbb{Z}_p(2)) := \ker(NH^3(X, \mathbb{Q}_p/\mathbb{Z}_p(2)) \to H^3(\overline{X}, \mathbb{Q}_p/\mathbb{Z}_p(2)))$$

By results of Bloch and Merkurjev-Suslin ([Bl], §5 and [M-S] we have the following exact sequence

$$(1-1) \qquad 0 \to H^1(X, \mathcal{K}_2) \otimes \mathbb{Q}_p/\mathbb{Z}_p \to NH^3(X, \mathbb{Q}_p/\mathbb{Z}_p(2)) \to CH_0(X)\{p\} \to 0$$

Since $H^1(\overline{X}, \mathcal{K}_2) \otimes \mathbb{Q}_p/\mathbb{Z}_p = 0$ we get an exact sequence

$$\begin{aligned}(1-2) \qquad & 0 \longrightarrow H^1(X, \mathcal{K}_2) \otimes \mathbb{Q}_p/\mathbb{Z}_p \longrightarrow K_N H^3(X, \mathbb{Q}_p/\mathbb{Z}_p(2)) \\ & \longrightarrow \ker(CH_0(X)\{p\} \longrightarrow CH_0(\overline{X})\{p\}^G) \longrightarrow 0\end{aligned}$$

Since X is identified with its Albanese variety, the map $CH_0(X)_{tors} \longrightarrow CH_0(\overline{X})^G_{tors}$ is the Albanese map and therefore $(CH_0(\overline{X})\{p\})^G \cong X(\mathbb{Q})\{p\}$ is finite. Consider the Hochschild-Serre spectral sequence

$$E_2^{a,b} = H^a(\mathbb{Q}, H^b(\overline{X}, \mathbb{Q}_p/\mathbb{Z}_p(2)) \Longrightarrow H^{a+b}(X, \mathbb{Q}_p/\mathbb{Z}_p(2)) \quad .$$

Then we have

LEMMA I: *Let the assumptions be as above. Then the composite map*

$$E_2^{2,1} \longrightarrow H^3(X, \mathbb{Q}_p/\mathbb{Z}_p(2)) \longrightarrow H^3(K, \mathbb{Q}_p/\mathbb{Z}_p(2))$$

is injective.

This is shown in ([L-S], Lemma (A)) without any assumption on the prime p.

COROLLARY (1.3) *The composite map*

$$\varphi : K_N H^3(X, \mathbb{Q}_p/\mathbb{Z}_p(2)) \longrightarrow H^1(G_{\mathbb{Q}}, A)$$

that is obtained by the Hochschild-Serre spectral sequence is injective.

The Corollary will play an important role in the proof of

LEMMA II: *Under the above assumptions let $p \geq 3$ be a prime such that $p \mid N$ and assume that the condition H 1) in the introduction is satisfied. Then we have*

$$H^1(X, \mathcal{K}_2) \otimes \mathbb{Q}_p/\mathbb{Z}_p = K_N H^3(X, \mathbb{Q}_p/\mathbb{Z}_p(2))_{\text{div}} \quad .$$

REMARK:

Lemma II was proven for primes $p \not| \, 6N$ in ([L-S, Lemma (B)) because in this case $K_N H^3(X, \mathbb{Q}_p/\mathbb{Z}_p(2))_{\text{div}}$ coincides with $H^1(\mathbb{Q}, A)_{\text{div}}$. This is not stated there explicitly but follows from the proof of Lemma (B) in [L-S].

Now we deduce Theorem A from Lemma II.

The exact sequence (1-1) also holds for a smooth proper model $\mathcal{X}$ of X over $\mathbb{Z}\left[\frac{1}{Np}\right]$. So $CH_0(\mathcal{X})\{p\}$ is a subquotient of $H^3(\mathcal{X}, \mathbb{Q}_p/\mathbb{Z}_p(2))$ and one knows that the latter group is co-finitely generated. Therefore $CH_0(\mathcal{X})\{p\}$ is co-finitely generated as $\mathbb{Z}_p$-module. Since the kernel of the canonical map

$$CH_0(\mathcal{X})\{p\} \longrightarrow CH_0(X)\{p\}$$

is a torsion group by the main result in [Mi], the localization sequence in the Zariski K-cohomology over $\mathcal{X}$ yields a surjection

$$CH_0(\mathcal{X})\{p\} \twoheadrightarrow CH_0(X)\{p\} \quad .$$

So we also know that $CH_0(X)\{p\}$ is co-finitely generated.

On the other hand, by (1-2), the finiteness of $CH_0(\overline{X})\{p\}^G$ and Lemma II we conclude that the maximal divisible subgroup of $CH_0(X)\{p\}$ is zero. Therefore $CH_0(X)\{p\}$ is a finite group.

To complete the proof of Theorem A it remains to show Lemma II.

§2

For each prime ℓ let

$$H^1_e(\mathbb{Q}_\ell, V) \subset H^1_f(\mathbb{Q}_\ell, V) \subset H^1_g(\mathbb{Q}_\ell, V) \subset H^1(\mathbb{Q}_\ell, V)$$

be defined as in ([BK], 3.7)). Let

$$H^1_f(\mathbb{Q}_\ell, T) \subset H^1_g(\mathbb{Q}_\ell, T) \subset H^1(\mathbb{Q}_\ell, T)$$

be the inverse image of $H^1_f(\mathbb{Q}_\ell, V)$ and $H^1_g(\mathbb{Q}_\ell, V)$. Put

$$H^1_f(\mathbb{Q}_\ell, A) := H^1_f(\mathbb{Q}_\ell, T) \otimes \mathbb{Q}_p/\mathbb{Z}_p \subset H^1(\mathbb{Q}_\ell, A)$$

and

$$H^1_g(\mathbb{Q}_\ell, A) := H^1_g(\mathbb{Q}_\ell, T) \otimes \mathbb{Q}_p/\mathbb{Z}_p \subset H^1(\mathbb{Q}_\ell, A)$$

Write $\wedge_\ell = H^1(\mathbb{Q}_\ell, T)/H^1_f(\mathbb{Q}_\ell, T)$. Then we have

$$\wedge_\ell \otimes \mathbb{Q}_p/\mathbb{Z}_p = H^1(\mathbb{Q}_\ell, A)_{\text{div}}/H^1_f(\mathbb{Q}_\ell, A)$$

Consider as in ([L-S], §3) the composite map

$$\psi : H^1(X, \mathcal{K}_2) \otimes \mathbb{Q}_p/\mathbb{Z}_p \longrightarrow K_N H^3(\overline{X}, \mathbb{Q}_p/\mathbb{Z}_p(2))_{\text{div}} \xrightarrow{\alpha'} \bigoplus_\ell \wedge_\ell \otimes \mathbb{Q}_p/\mathbb{Z}_p$$

where α' is the restriction of the map

$$\alpha : H^1(\mathbb{Q}, A) \longrightarrow \bigoplus_{all\ \ell} \frac{H^1(\mathbb{Q}_\ell, A)}{H^1_f(\mathbb{Q}_\ell, A)}$$

the kernel of which defines the Selmer group $S(\mathbb{Q}, A)$.

In analogy to ([L-S], Lemma 3.1) we will prove the following

PROPOSITION (2.1): *Let the notations be as in §1. Let $p \geq 3$ a prime, such that E has multiplicative reduction at p. Assume that condition (H 1) holds. Then we have*

a) $$\operatorname{coker}\psi = H^1(\mathbb{Q}_p, A)_{\mathrm{div}}/H^1_g(\mathbb{Q}_p, A)$$

b) $$\operatorname{Im}\psi = \operatorname{Im}\alpha'$$

We will give the proof of Proposition 2.1 in the next section.

In the following we will compute the coranks of $H^1(\mathbb{Q}_p, A)_{\mathrm{div}}/H^1_g(\mathbb{Q}_p, A)$ and $H^1_g(\mathbb{Q}_p, A)/H^1_f(\mathbb{Q}_p, A)$. Let

$$\Omega_p = H^1_g(\mathbb{Q}_p, V)/H^1_f(\mathbb{Q}_p, V) \text{ and } \theta_p = H^1(\mathbb{Q}_p, V)/H^1_g(\mathbb{Q}_p, V)$$

as in ([L-S], §4). It is well known that $X_{\mathbb{Q}_p} = E \times E_{\mathbb{Q}_p}$ has a regular proper model $\mathcal{X}$ over $\mathbb{Z}_p$ with semistable reduction. Let X_p be its closed fiber. By local Tate-Duality ([B-K], §3.8), Ω_p is the $\mathbb{Q}_p$-dual of $H^1_f(\mathbb{Q}_p, V(-1))/H^1_e(\mathbb{Q}_p, V(-1))$ and this quotient is — by the computations in [B-K], 3.8 — isomorphic to $(B_{\mathrm{crys}} \otimes V(-1))^{G_{\mathbb{Q}_p}}/1-f$, which is by Kato's and Tsuji's proof of the Fontaine-Jannsen-Conjecture ([Ka], §6), ([Tsu]) isomorphic to $(D_2)^{N=0}/1-f)$, where

$$D_2 = H^2_{\log\ \mathrm{crys}}((X_p, M_1)/W(\mathbb{F}_p), W(L), O^{\mathrm{crys}}) \otimes \mathbb{Q}_p$$

denotes the log-crystalline cohomology introduced by Hyodo-Kato [H-K], $N = 0$ denotes the kernel under the action of the monodromy operator N, and f acts as $p^{-1}\varphi$, where φ is the Frobenius acting on D_2. Therefore we have by Poincaré duality for Hyodo-Kato cohomology that Ω_p is isomorphic to $(\operatorname{coker} N : D_2 \to D_2)^{\varphi=p}$. Since the functor $D_{st}(\cdot) = (B_{st} \otimes \cdot)^{G_{\mathbb{Q}_p}}$ commutes with tensor products and a Tate-elliptic curve has ordinary semistable reduction in the sense of ([Il], Definition 1.4) we have a Hodge-Witt-decomposition ([Il], Proposition 1.5)

$$D_2 = \bigoplus_{i+j=2} H^i(X_p, Ww^j) \otimes \mathbb{Q}_p \quad .$$

Here $H^i(X_p, Ww^j)$ is the cohomology of the modified Hodge-Witt-sheaves. From the action of the Frobenius φ on D_2 it is clear that $(D_2)_{\varphi=p}$ is contained in $H^1(X_p, Ww^1)_{\mathbb{Q}_p}$. By ([Mo], §6) we know that the monodromy filtration and the weight filtration on D_2 coincide. Using the formula $N\varphi = p\varphi N$ we have that

$$N(H^0(X_p, Ww^2)) \subset H^1(X_p, Ww^1)$$

and the map

$$N^2 : H^0(X_p, Ww^2) \longrightarrow H^2(X_p, Ww^0)$$

is an isomorphism. Since $\dim H^i(X_p, Ww^j)_{\mathbb{Q}_p} = \dim H^i(X_{\mathbb{Q}_p}\Omega^j)$ by ([Il], Corollaire 2.6), we see that

$$\dim(\operatorname{coker} N : D_2 \to D_2)^{\varphi=p} = \dim(D_2)^{N=0}_{\varphi=p} \leq 3 \quad .$$

On the other hand the B_{St}-comparison-isomorphism provides an injection

$$\operatorname{Pic}(X) \otimes \mathbb{Q}_p \hookrightarrow H^2(\overline{X}, \mathbb{Q}_p(1))^{G_{\mathbb{Q}_p}} \hookrightarrow (D_2)^{N=0}_{\varphi=p} \quad .$$

Since $\operatorname{Pic}(X)$ has rank 3 we have

Lemma (2.2):

$$\dim \Omega_p = \dim (D_2)^{N=0}_{\varphi=p} = 3 \quad .$$

By the same methods and the proof of ([L-S], Lemma 4.4) we get

Lemma (2.3):

$$\dim \theta_p = 1 \quad .$$

From Lemma (2.2) and ([L-S], Lemma 4.1) we get

Lemma (2.4): *The image of the composite map*

$$(\mathrm{Pic}(X) \otimes \mathbb{Q}^*) \otimes \mathbb{Q}_p/\mathbb{Z}_p \longrightarrow H^1(X, \mathcal{K}_2) \otimes \mathbb{Q}_p/\mathbb{Z}_p \xrightarrow{\psi_p} \wedge_p \otimes \mathbb{Q}_p/\mathbb{Z}_p$$

is

$$H^1_g(\mathbb{Q}_p, A)/H^1_f(\mathbb{Q}_p, A) \quad .$$

Now we will give the proof of Theorem B and we distinguish between two cases.

Case I:

The map α'_p, i.e. the p-component of α' is surjective.

This case is actually obstructed by the Gersten-conjecture as we will see in the proof of Proposition (2.1). Since we do not assume (H 1) in Theorem B we also consider this case. Using the surjectivity-property of ψ_ℓ, i.e. the ℓ-component of ψ, for $\ell \neq p$ that follows from Prop. 2.1, and where the condition (H 1) is not needed, we see that $\operatorname{coker} \alpha$ has $\mathbb{Z}_p$-corank 0. Now apply the modified version of ([B-K], Lemma 5.16) that is given in ([L-S], Lemma (3.3)): All the assumptions there are also satisfied for our choice of p:

— V is a de Rham representation of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ by Falting's proof of the de Rham conjecture.

— For the characteristic polynomial $P_\ell(V,t)$ we have $P_\ell(V,1) \neq 0$. For $\ell \neq p$ the proof is the same as in ([L-S], §3). For $\ell = p$, we have $\mathrm{Crys}(V)^{f=1} = (D_2)^{N=0}_{\varphi=p^2}$. By the same methods as in the proof of Lemma (2.2) we have $(D_2)^{N=0}_{\varphi=p^2} = 0$.
By the same arguments as in the proof of ([L-S], Theorem 3.2) we get the formula $\operatorname{corank}(\ker \alpha) = \operatorname{corank}(\operatorname{coker} \alpha) = 0$. Therefore $S(\mathbb{Q}, A) = \ker \alpha$ is finite.

Case II:

$\operatorname{Im} \alpha'_p = H^1_g(\mathbb{Q}_p, A)/H^1_f(\mathbb{Q}_p, A)$

By Lemmas (2.3) and (2.4) this is the only remaining case to consider.

Let $T' = \mathrm{Sym}^2 H^1(\overline{E}, \mathbb{Z}_p(1))$. By Lemma (2.2) and Lemma (2.4) we have $H^1_g(\mathbb{Q}_p, T')/H^1_f(\mathbb{Q}_p, T') = 0$. Let $c(\ell)$ for $\ell \not| N$ be the elements in $H^1(X, \mathcal{K}_2)$ that were constructed by Mildenhall and Flach. In the notation of ([Fl], Prop. (1.1)) we therefore have $\mathrm{res}_{r=p}c(\ell) \in H^1_f(\mathbb{Q}_p, T')$. We get this property with little effort whereas in [Fl] this was one of the harder parts in the whole paper. It is now easy to check that all the other required properties on the elements $c(\ell)$ in ([Fl], Prop. (1.1)) are also satisfied for our choice of p. Thus we apply Kolyvagin's argument in ([Fl], Prop. (1.1)). At the point where Flach needs the surjectivity of the Galois representation ϱ_p in order to derive the finiteness of $S(\mathbb{Q}, A(-1))$, we use the following Lemma, due to Nekovář, that finishes, after applying Poitou-Tate Duality, the proof of Theorem B.

LEMMA (2.5): *Let $\mathbb{Q}(E_{p^n})/\mathbb{Q}$ be the Galois extension obtained by adjoining the coordinates of all p^n-torsion points on E and let T' be as above. Then there exists a $c > 0$, such that the exponent of $H^1(\mathrm{Gal}(\mathbb{Q}(E_{p^n})/\mathbb{Q}), T'(-1)/p^n)$ divides p^c for all $n \geq 0$.*

REMARK: Flach uses the vanishing of this cohomology group that follows from his additional assumption on the surjectivity of ϱ_p.

PROOF: Put $G := \mathrm{Im}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}_{\mathbb{Z}_p}(T_p(E)))$. Since E is without complex multiplication over $\overline{\mathbb{Q}}$, G is of finite index in $\mathrm{Aut}_{\mathbb{Z}_p}(T_p(E)) = GL_2(\mathbb{Z}_p)$. Put $G_n := \ker(G \to GL_2(\mathbb{Z}/p^n), T' := \mathrm{Sym}^2(T_p(E))$, $\tilde{G} := \mathrm{Im}(G \to \mathrm{Aut}_{\mathbb{Z}_p}(T')) = G/Z \cap G$, where $Z = \text{center of } GL_2(\mathbb{Z}_p) = \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, \lambda \in \mathbb{Z}_p^* \right\}$.
Consider the following diagram with horizontal and vertical exact sequences:
(note that $G/G_n \cong \mathrm{Gal}(\mathbb{Q}(E_{p^n})/\mathbb{Q})$.

$$\begin{array}{ccccc}
 & & 0 & & \\
 & & \downarrow & & \\
 & & H^1(G, T'(-1)) \otimes \mathbb{Z}/p^n & & \\
 & & \downarrow & & \\
0 \to H^1(G/G_n, T'(-1)/p^n) & \xrightarrow{\mathrm{inf}} & H^1(G, T'(-1)/p^n) & \xrightarrow{\mathrm{res}} & H^1(G_n, T'(-1)/p^n)^{G/G_n} \\
 & & \downarrow & & \\
 & & H^2(G, T'(-1))_{p^n} & &
\end{array}$$

It is clear that $H^i(G, T'(-1)) = H^i_{\mathrm{cont}}(G, T'(-1)) = H^i_{\mathrm{naive}}(G, T'(-1))$ are $\mathbb{Z}_p$-modules of finite type. Therefore $H^2(G, T'(-1))_{p^\infty}$ is finite. We have an exact sequence

$$\begin{array}{l}
0 \to H^1(\tilde{G}, T'(-1)) \xrightarrow{\mathrm{inf}} H^1(G, T'(-1)) \xrightarrow{\mathrm{res}} H^1(Z \cap G, T'(-1))^{G/Z\cap G} \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = \\
\qquad\qquad\qquad\qquad\qquad\qquad \mathrm{Hom}_{\mathrm{cont}}(Z \cap G, (T'(-1))^{G/Z\cap G})
\end{array}$$

But $(T'(-1))^{G/Z\cap G}$ is zero (E has no CM). Thus $H^1(\tilde{G}, T'(-1)) = H^1(G, T'(-1))$. By result of Lazard there is an injection

$$\begin{aligned} H^1(\tilde{G}, T'(-1)) \otimes \mathbb{Q} \quad \hookrightarrow \quad & H^1(Lie(\tilde{G}), T'(-1) \otimes \mathbb{Q}) \\ & = H^1(sl(2), T'(-1) \otimes \mathbb{Q}) \end{aligned}$$

and H^1 vanishes for semisimple Lie-algebras (and every representation). So $H^1(G, T'(-1))$ is finite and Lemma 2.5 follows.

Finally it is easy to see that Corollary (1.3), Proposition (2.1) b) and Theorem B imply Lemma II and as a consequence also Theorem A. It remains to show Proposition (2.1). This will be accomplished in the next paragraph.

§3

The surjectivity of the map

$$\psi' = \bigoplus_{\ell \neq p} \psi_\ell : H^1(X, \mathcal{K}_2) \otimes \mathbb{Q}_p/\mathbb{Z}_p \longrightarrow \bigoplus_{\ell \neq p} H^1(\mathbb{Q}_\ell, A)_{\mathrm{div}}/H^1_f(\mathbb{Q}_\ell, A)$$

follows from ([L-S], Lemmas (4.1), (4.3), (4.4) and (4.5)). On the other hand the composite map

$$\mathrm{Pic}(X) \otimes p^{\mathbb{Z}}) \otimes \mathbb{Q}_p/\mathbb{Z}_p \to H^1(X, \mathcal{K}_2) \otimes \mathbb{Q}_p/\mathbb{Z}_p \stackrel{\psi_p}{\to} H^1_g(\mathbb{Q}_p, A)/H^1_f(\mathbb{Q}_p, A)$$

is surjective by Lemma (2.2), whereas the image of $(\mathrm{Pic}(X) \otimes p^{\mathbb{Z}}) \otimes \mathbb{Q}_p/\mathbb{Z}_p$ under the map ψ' is zero. To finish the proof of Proposition (2.1) we therefore have to show that the image of α'_p, the p-component of α' is contained in $H^1_g(\mathbb{Q}_p, A)/H^1_f(\mathbb{Q}_p, A)$.

By the theory of Bloch-Ogus and the work of Merkurjev-Suslin [M-S] we have an isomorphism

$$H^1(X, \mathcal{K}_2/p^n) \cong NH^3_{et}(X, \mathbb{Z}/p^n(2)) \quad .$$

Let $\mathcal{X}$ be a proper regular semistable model of $X_{\mathbb{Q}_p}$ over $\mathbb{Z}_p$, $i : X_p \to \mathcal{X}$ and $j : X_{\mathbb{Q}_p} \hookrightarrow \mathcal{X}$ the inclusions of the closed and generic fiber.

Let $H^3_{et}(\mathcal{X}, \tau_{\leq 2} Rj_* \mathbb{Z}/p^n(2))$ be the cohomology of the truncated complex of p-adic vanishing cycles. Then we have

LEMMA (3.1): *Assume that the Gersten-Conjecture holds for the Zariski sheaf $\mathcal{K}_2$ on the regular scheme $\mathcal{X}$. Then we have the inclusion*

$$H^1(X_{\mathbb{Q}_p}, \mathcal{K}_2/p^n) \subset H^3_{et}(\mathcal{X}, \tau_{\leq 2} Rj_* \mathbb{Z}/p^n(2)) \quad .$$

PROOF:

This follows from the proof of ([L-S], Lemma (5.4)).

LEMMA (3.2): $H^3(\overline{X}_{\mathbb{Q}_p}, \mathbb{Q}_p(2))^{G_{\mathbb{Q}_p}} = 0$.

PROOF:

Using the Künneth formula and the fact that $H^2(\overline{E}, \mathbb{Q}_p(1)) \cong \mathbb{Q}_p$ (the Brauer group of a curve over an algebraically closed field is zero), it suffices to show that $H^1(\overline{E}, \mathbb{Q}_p(1))^{G_{\mathbb{Q}_p}} = 0$. This follows from ([J], Theorem 5a).

Using Lemma (3.2) and the Hochschild-Serre spectral sequence we get a canonical map

$$\sigma : \varprojlim_n H^1(X_{\mathbb{Q}_p}, \mathcal{K}_2/p^n) \otimes \mathbb{Q}_p \longrightarrow H^1(\mathbb{Q}_p, V) \quad .$$

When we deal with a variety over a local field, all cohomology groups under consideration are (co-)finitely generated. The map α'_p certainly factors through $\varinjlim_n H^1(X_{\mathbb{Q}_p}, \mathcal{K}_2/p^n)_{\mathrm{div}}$. The assertion that $\varinjlim_n H^1(X_{\mathbb{Q}_p}, \mathcal{K}_2/p^n)_{\mathrm{div}}$ is contained in $H^1_g(\mathbb{Q}_p, A)$ is therefore equivalent to the assertion that the image of σ is contained in $H^1_g(\mathbb{Q}_p, V)$. In view of Lemma (3.1) we see that Proposition (2.1) follows from the following

LEMMA (3.3): *Under the condition H1) we have:* $\mathrm{Im}\,\sigma \subset H^1_g(\mathbb{Q}_p, V)$.

To prove Lemma (3.3) it suffices to show that the image of the map

$$H^3(\mathcal{X}, \tau_{\leq 2} Rj_* \mathbb{Q}_p(2)) \longrightarrow H^1(\mathbb{Q}_p, V)$$

is contained in $H^1_g(\mathbb{Q}_p, V)$.

Let $s_n^{\log}(2)$ be the log-syntomic complex in $D_{et}(\mathcal{X})$ constructed by Kato ([Ka], §6) and Tsuji [Tsu] together with a canonical map

$$s_n^{\log}(2) \longrightarrow \tau_{\leq 2} i_* i^* Rj_* \mathbb{Z}/p^n(2) \quad .$$

This gives rise to a composite map

$$\eta : H^3_{et}(\mathcal{X}, s^{\log}_{\mathbb{Q}_p}(2)) \longrightarrow H^1(\mathbb{Q}_p, V) \quad .$$

Since $(D_2)^{N=0}_{\varphi=p^2} = (D_3)^{N=0}_{\varphi=p^2} = 0$ (D_i denotes the i-th log-crystalline cohomology of X_p) we may apply the main result in [L] on a semistable analogue of Schneider's p-adic points conjecture to get

LEMMA (3.4) $\operatorname{Im}\eta = H^1_g(\mathbb{Q}_p, V)$.

Tsuji has proven that there is a canonical isomorphism between the cohomology $\mathcal{H}^2(i^* s_n^{\log}(2))$ and the sheaf $M_n^2 = i^* R^2 j_* \mathbb{Z}/p^n(2)$ of p-adic vanishing cycles ([Tsu], Theorem 3.2). His proof relies on a filtration $\mathrm{Fil}^{\cdot}$ on M_n^2 that was defined by Hyodo ([H], (1.4)) and is induced by a symbol map on Milnor K-Theory. Hyodo has shown ([H], Theorem (1.6)) that the highest graded quotient $gr^0 M_n^2$ sits in an extension (change of notation: $Y := X_p$, the closed fiber of $\mathcal{X}$)

$$0 \longrightarrow W_n w^1_{Y,log} \longrightarrow gr^0 M_n^2 \longrightarrow W_n w^2_{Y,log} \longrightarrow 0$$

where $W_n w^i_{Y,log}$ are the modified logarithmic Hodge-Witt-sheaves ([H] (1.5)). On the other hand Hyodo and Kato ([H-K] Prop. 1.5) constructed an exact sequence of Hodge-Witt-sheaves

$$0 \longrightarrow W_n w^1_Y \longrightarrow W_n \tilde{w}^2_Y \longrightarrow W_n w^2_Y \longrightarrow 0$$

and used the connecting homomorphism on the level of cohomology to define the monodromy operator on log-crystalline cohomology. It follows from the work of Tsuji ([Tsu], §2.4) that there is a commutative diagram

$$\begin{array}{ccccccccc} 0 & \to & W_n w^1_{Y,log} & \to & gr^0 M_n^2 & \to & W_n w^2_{Y,log} & \to & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \to & W_n w^1_Y & \to & W_n \tilde{w}^2_Y & \to & W_n w^2_Y & \to & 0 \end{array}$$

such that the upper exact sequence is obtained by taking the kernel of $1 - F$ acting on the lower exact sequence, where F is the Frobenius. From the Hodge-Witt-decomposition of $H^r(Y, Ww^{\cdot})$ ([Il], Proposition (1.5)) it is easy to derive a Hodge-Witt-decomposition for $H^r(Y, W\tilde{w}^{\cdot}_Y)$

$$H^r(Y, W\tilde{w}^{\cdot}_Y) = \bigoplus_{i+j=r} H^i(Y, W\tilde{w}^j_Y) \quad .$$

From the action of the Frobenius φ on $H^r(Y, W\tilde{w}^{\cdot}_Y)$ we get

$$H^3(Y, W\tilde{w}^{\cdot}_Y)_{\varphi=p^2} = H^1(Y, W\tilde{w}^2_Y)^{F=1} \quad .$$

On the other hand it is shown in the proof of the semistable analogue of the p-adic points conjecture on log-syntomic cohomology [L], (2.6), Prop. (2.9), Prop. (2.13) that there is a surjection

$$H^3_{et}(\mathcal{X}, s^{\log}_{\mathbb{Q}_p}(2)) \twoheadrightarrow (H^3(Y, W\tilde{w}^{\cdot}_Y)_{\mathbb{Q}_p})_{\varphi=p^2}$$

and the above arguments yield a commutative diagram

$$\begin{array}{ccccc}
H^3(\mathcal{X}, s^{\log}_{\mathbb{Q}_p}(2)) & & & & \\
\downarrow & \searrow & & & \\
H^3_{et}(\mathcal{X}, \tau_{\leq 2} Rj_* \mathbb{Q}_p(2)) & \to & H^1(Y, gr^0 M^2_{\mathbb{Q}_p}) & \to & (H^3(Y, W\tilde{\omega}^{\cdot}_Y)_{\mathbb{Q}_p})_{\varphi=p^2} \\
\downarrow & & & & \downarrow \\
H^1(\mathbb{Q}_p, V) & & \longrightarrow & & H^1(\mathbb{Q}_p, B_{\mathrm{crys}} \otimes V)
\end{array}$$

It follows from ([L], (2.10)) that the composite

$$(H^3(Y, W\tilde{\omega}^{\cdot}_Y)_{\mathbb{Q}_p})_{\varphi=p^2} \longrightarrow H^1(\mathbb{Q}_p, B_{\mathrm{crys}} \otimes V) \longrightarrow H^1(\mathbb{Q}_p, B_{st} \otimes V)$$

is the zero map. Using the fact that $H^1_{st} = H^1_g$ (unpublished result of Hyodo, see also Nekovář ([Ne](1.24)) we conclude that the image of the map

$$H^3_{et}(\mathcal{X}, \tau_{\leq 2} Rj_* \mathbb{Q}_p(2)) \longrightarrow H^1(\mathbb{Q}_p, V)$$

is $H^1_g(\mathbb{Q}_p, V)$ in view of Lemma (3.4). This finishes the proof of Lemma (3.3) and Proposition (2.1).

REFERENCES

[Bl] S. Bloch, *Lectures on Algebraic Cycles.* Duke Univ. Math. Series, Durham, 1980.

[B-K] S. Bloch, K. Kato, *L-functions and Tamagawa numbers of motives.* In: The Grothendieck Festschrift 1990.

[Fl] M. Flach, *A finiteness theorem of the symmetric square of an elliptic curve.* Invent. Math. 109 (1992), 307–327.

[H] O. Hyodo, *A note on p-adic étale cohomology in the semistable reduction case.* Invent. Math. 91 (1988), 543–557.

[H-K] O. Hyodo, K. Kato, *Semistable reduction and crystalline cohomology with logarithmic poles.* In: Périodes p-adiques, Astérisque 223 (1994), 221–268.

[Il] L. Illusie, *Réduction semistable ordinaire, cohomologie étale p-adique et cohomologie de de Rham d'après Bloch-Kato et Hyodo.* In: Périodes p-adiques, Astérisque 223 (1994), 209–220.

[J] U. Jannsen, *On the ℓ-adic cohomology of varieties over number fields and its Galois-cohomology.* In: Galois groups over $\mathbb{Q}$, MSRI Publications (1989), 315–360.

[K] K. Kato, *Semistable reduction and p-adic etal cohomology.* In: Périodes p-adiques, Astérisque 223 (1994), 269–293.

[L] A. Langer, *Local points of motives in semistable reduction.* To appear in Compositio Mathematica.

[L-S] A. Langer, S. Saito, *Torsion zero-cycles on the selfproduct of a modular elliptic curve.* In: Duke Math. Journal, November 1996.

[Mi] S. Mildenhall, *Cycles in a product of elliptic curves and a group analogous to the class group.* Duke Math. J. 67 (1992), 387–406.

[Mo] A. Mokrane, *La suite spectrale des poids en cohomologie de Hyodo-Kato.* Duke Math. J. 72 (1993), 301–337.

[M-S] A. S. Merkurjev, A. A. Suslin, *K-cohomology of Severi-Brauer. Varieties and the norm residue homomorphism.* Math. USSR Izvestiya 21, No. 2 (1983), 307–340.

[Ne] J. Nekovář, *On p-adic height-pairings.* In: Seminaire de Théorie de Nombres 1990–1991 (1993), 127–202.

[Tsu] T. Tsuji, *p-adic etale cohomology and crystalline cohomology in the semi-stable reduction case.* Preprint 1996.

[W] A. Wiles, *Modular elliptic curves and Fermat's last Theorem.* Annals of Mathematics 142 (1995), 443–551.

Andreas Langer
Mathematisches Institut
der Universität Münster
Einsteinstr. 62
D - 48149 Münster
Germany

Hopf-Bifurcation in Systems with Spherical Symmetry Part I : Invariant Tori

Christian Leis

Received: February 10, 1997
Revised: April 4, 1997

Communicated by Bernold Fiedler

Abstract.

A Hopf-bifurcation scenario with symmetries is studied. Here, apart from the well known branches of periodic solutions, other bifurcation phenomena have to occur as it is shown in the second part of the paper using topological arguments. In this first part of the paper we prove analytically that invariant tori with quasiperiodic motion bifurcate. The main methods used are orbit space reduction and singular perturbation theory.

1991 Mathematics Subject Classification: 58F14, 34C20, 57S15

Contents

1 Introduction

An interesting problem in the theory of ordinary differential equations is the generalization of the two dimensional Hopf-bifurcation to higher dimensional systems with symmetry. In this connection, [GoSt] and [GoStSch] investigated problems on a vector space X that can be decomposed into a direct sum of absolutely irreducible representations of the group O(3) of the form $X = V_l \oplus \mathrm{i}V_l$. Here V_l denotes the space of homogeneous harmonic polynomials $P : \mathbb{R}^3 \to \mathbb{R}$ of degree l. This is the simplest case where purely imaginary eigenvalues (of high multiplicity) in the bifurcation point are possible. Using Lie-group theory, the authors showed the existence of branches of periodic solutions with certain symmetries. Here in addition to the spatial O(3)-symmetry a temporal S^1-symmetry occurs. This symmetry corresponds to a time shift along the periodic solutions. In order to obtain their results, the authors made a Lyapunov-Schmidt-reduction on the space of periodic functions. The reduced system then has $\mathrm{O}(3) \times \mathrm{S}^1$-symmetry and solutions correspond to periodic solutions of the original system with spatial-temporal symmetry. Under certain transversality assumptions, periodic solutions with symmetry $\bar{H} \subset \mathrm{O}(3) \times \mathrm{S}^1$ bifurcate if Dim Fix$(\bar{H}) = 2$ for the induced representation of the group $\mathrm{O}(3) \times \mathrm{S}^1$ on the space X (cf. [GoSt] resp. [GoStSch]). [Fi] has shown that it is sufficient that $\bar{H}$ is a maximal subgroup for periodic solutions with symmetry $\bar{H}$ to bifurcate. Using these methods, only the existence of periodic solutions can be investigated. Via normal form theory (cf. [EletAl]) one gets $\mathrm{O}(3) \times \mathrm{S}^1$-equivariant polynomial vector fields up to every finite order for our systems. This additional S^1-symmetry is due to the fact that the normal form commutes with the one parameter group $e^{L^{\mathrm{T}}t}$ which is generated by the linearization L in the bifurcation point. For a Hopf-bifurcation L has purely imaginary eigenvalues (of high multiplicity) and the group generated is a rotation. [IoRo], [HaRoSt] and [MoRoSt] did analytic calculations for the normal form up to fifth order in the case $l = 2$. They gave conditions for the stability of the five branches of periodic solutions predicted by [GoSt] resp. [GoStSch] in terms of coefficients of the normal form. Quasiperiodic solutions found by [IoRo] in the normal form up to third order can not be confirmed in this paper. We shall show a mechanism for quasiperiodic solutions to bifurcate in the fifth order.
Investigating the normal form due to [IoRo], one finds a region in parameter space where two of the branches of periodic solutions bifurcating supercritically are stable simultaneously. Using topological methods, [Le] showed that we have the following alternative in this region in parameter space: Either besides the known branches of periodic solutions other invariant objects bifurcate or recurrent structure between the different invariant sets (e.g. between the different group orbits of periodic solutions and the trivial solution) exists. Actually the results of these topological investigations were the starting point of analytical efforts to find other solutions (or recurrent structure) in this paper. In order to get our results, we shall proceed as follows.
First the representation of the group $\Gamma = \mathrm{O}(3) \times \mathrm{S}^1$ on the ten dimensional space $X = V_2 \oplus \mathrm{i}V_2$ is introduced. The lattice of isotropy subgroups of this representation is given according to [MoRoSt] and the results of [IoRo] are quoted. The smallest invariant subspace containing both solutions that are stable simultaneously has isotropy $\Sigma = (\mathbb{Z}_2, 1)$.

Then our considerations are being restricted to this six dimensional subspace. The normaliser of Σ is $\mathrm{N}(\Sigma) = \mathrm{O}(2) \times \mathrm{S}^1 \subset \Gamma$. This is the biggest subgroup of Γ leaving $\mathrm{Fix}(\Sigma)$ invariant as a subspace. Now we shall look at the representation of $\frac{\mathrm{N}(\Sigma)}{\Sigma}$ on $\mathrm{Fix}(\Sigma)$.

Dealing with differential equations with symmetries, one has to deal with group orbits of solutions because a solution $x(t)$ gives rise to solutions $\gamma\, x(t)$ with $\gamma \in \Gamma$. This redundancy, induced by the action of the group, will be removed by identifying points that lie on a group orbit. I.e. one studies the orbit space that is homeomorphic to the image of the Hilbert-map $\Pi : \mathrm{Fix}(\Sigma) \to \mathbb{R}^k : z \to \pi_i(z)$ (cf. [La2] and [Bi]). Here k denotes the minimal number of generators of the ring of $\frac{\mathrm{N}(\Sigma)}{\Sigma}$ invariant polynomials $P : \mathrm{Fix}(\Sigma) \to \mathbb{R}$ and π_i, $i = 1, \ldots, k$, is such a system of generators. Thus the original differential equation is reduced to a differential equation on $\Pi\big(\mathrm{Fix}(\Sigma)\big)$ of the form $\dot{\pi} = g(\pi)$, $\pi = (\pi_1, \ldots, \pi_k)$. In order to perform this reduction for a given equation, one, first of all, has to know the number of independent invariants and equivariants for a given representation. Then one, actually, has to calculate them. Statements on the number of independent invariants and equivariants and possible relations between them are given by the Poincaré-series. These are formal power series $\sum_{i=0}^{\infty} a_i\, t^i$ in t. Here a_i denotes the dimension of the vector space of homogeneous invariant polynomials of degree i resp. the dimension of the vector space of homogeneous equivariant mappings of degree i. These series can be determined just by knowledge of the representation of the group on the space.
The lattice of isotropy subgroups of the representation of $\frac{\mathrm{N}(\Sigma)}{\Sigma}$ on $\mathrm{Fix}(\Sigma)$ and the image of the Hilbert-map are determined. This is a stratified space which consists of manifolds (strata). Each stratum consists of images of points of some isotropy type of the representation of $\frac{\mathrm{N}(\Sigma)}{\Sigma}$ on $\mathrm{Fix}(\Sigma)$. Thus it is flow invariant with respect to the reduced vector field on $\Pi\big(\mathrm{Fix}(\Sigma)\big)$.
Afterwards we shall carry out the orbit space reduction for the normal form up to third order. The critical points of the reduced vector field in $\Pi\big(\mathrm{Fix}(\Sigma)\big)$ are determined. As expected by inspection of the lattice of isotropy subgroups of Γ on $V_2 \oplus \mathrm{i}V_2$, we shall find images of periodic solutions of isotropy $(\mathrm{O}(2), 1)$, $(\mathrm{D}_4, \mathbb{Z}_2)$, $\widetilde{\mathrm{SO}(2)}^2$, and $(\mathbb{T}, \mathbb{Z}_3)$. Moreover there exists some stratum F in $\Pi\big(\mathrm{Fix}(\Sigma)\big)$. Connected via a curve g of fixed points the fixed points having isotropy $(\mathrm{O}(2), 1)$ resp. $(\mathrm{D}_4, \mathbb{Z}_2)$ in the original system lie on F. The preimage of F consists of points having isotropy $(\mathbb{Z}_2, 1)$ in the restricted system. Perturbations that respect the symmetry will, therefore, respect this stratum. The curve g is stable for the reduced vector field restricted to F. Small perturbations of the original vector field in fifth order of magnitude ε will, therefore, preserve a curve. By use of singular perturbation theory (cf. [Fe]), one gets a resulting drift on the curve. This explains the observation made by [IoRo] that the stability of the fixed points of isotropy $(\mathrm{O}(2), 1)$ resp. $(\mathrm{D}_2, \mathbb{Z}_2)$ is determined in the fifth order.
Dependent on the relative choice of the coefficients of the third order normal form in the region of parameter space in question, there is a point on the curve g where the linear stability of the curve in the direction of the principle stratum changes. Linearization of the reduced vector field in this point yields a nontrivial two dimensional Jordan-block to the eigenvalue zero. The second dimension results from the

linearization along the curve. Finally the flow on the two dimensional center manifold in this point is determined for small ε. The persistence of the curve g for small ε, knowledge of the direction of the drift, the change of stability in the direction of the principle stratum, and the existence of a nontrivial two dimensional Jordan-block to the eigenvalue zero are sufficient to prove for small ε the bifurcation of a fixed point of the reduced equation in the direction of the principle stratum using the implicit function theorem. Fixed points of the reduced system on the stratum F correspond to periodic solutions, fixed points in the principle stratum correspond to quasiperiodic solutions in the original system.

2 Representation of the group $\mathrm{O}(3) \times \mathrm{S}^1$ on $V_2 \oplus \mathrm{i}V_2$

We investigate systems of ODE's of the form

$$\dot{x} \quad = \quad f(\lambda, x)$$

in the ten dimensional space

$$X \quad = \quad V_2 \oplus \mathrm{i}V_2.$$

Let V_2 be the five dimensional space of homogeneous harmonic polynomials

$$p \quad : \quad \mathbb{R}^3 \quad \to \quad \mathbb{R}$$

of degree two. We have

$$V_2 \quad = \quad \left\langle 2x_3^2 - (x_1^2 + x_2^2), x_1x_3, x_2x_3, x_1^2 - x_2^2, x_1x_2 \right\rangle .$$

Let us introduce the following coordinates $(z, \overline{z})$,

$$z \quad = \quad (z_{-2}, z_{-1}, z_0, z_1, z_2), \;\; z_m \in \mathbb{C}, \;\; m = -2, \ldots, 2,$$

in the space X:

$$x \in X \quad \Leftrightarrow \quad x = \sum_{m=-2}^{2} z_m \, Y_m .$$

Here

$$\begin{aligned} Y_0 &= \sqrt{\tfrac{5}{16\pi}} \left(2x_3^2 - (x_1^2 + x_2^2)\right), \\ Y_{\pm 1} &= \sqrt{\tfrac{15}{8\pi}} \left(x_1x_3 \pm \mathrm{i}x_2x_3\right), \\ Y_{\pm 2} &= \sqrt{\tfrac{15}{32\pi}} \left((x_1^2 - x_2^2) \pm \mathrm{i}2x_1x_2\right) \end{aligned}$$

denote spherical harmonics. Moreover let

$$f \quad : \quad \mathbb{R} \times X \quad \to \quad X$$

be a smooth map that commutes with the following representation of the compact Lie-group

$$\Gamma \quad = \quad \mathrm{O}(3) \times \mathrm{S}^1$$

on the space X.
The group

$$\mathrm{O}(3) = \mathrm{SO}(3) \oplus \mathbb{Z}_2^c$$

with

$$\mathbb{Z}_2^c = \{\pm \mathrm{Id}\}$$

acts via the natural representation absolutely irreducible on V_2. For $p \in V_2$ and $\gamma \in \Gamma$ we have

$$\begin{aligned} \gamma\, p(\cdot) &= p(\gamma^{-1}\cdot) \text{ for } \gamma \in \mathrm{SO}(3), \\ -\mathrm{Id}\, p(\cdot) &= p(\cdot). \end{aligned}$$

This representation is a special case of the representation of the group O(3) on the space V_l, $l \geq 1$. For l even the subgroup $\mathbb{Z}_2^c$ acts trivially in the natural representation. On the space X the group O(3) acts diagonally. For the general representation theory of O(3) we refer to [StiFä] and [GoStSch].
The group S^1 acts as a rotation in the coordinates

$$\begin{aligned} \phi\, z &= e^{i\phi} z, \\ \phi\, \overline{z} &= e^{-i\phi} \overline{z} \end{aligned}$$

with $\phi \in \mathrm{S}^1$.
So we have

$$f(\lambda, \gamma\, x) = \gamma\, f(\lambda, x), \ \forall \gamma \in \Gamma.$$

In their paper concerning Hopf-bifurcation with O(3)-Symmetry [GoSt] and [GoStSch] look at systems of the form

$$\dot{x} = f(\lambda, x)$$

with

$$x \in X = V_l \oplus \mathrm{i} V_l$$

and

$$f \ : \ \mathbb{R} \times X \ \to \ X$$

a smooth mapping. This direct sum of two absolutely irreducible representations of the group O(3) is the simplest case allowing imaginary eigenvalues, however of high multiplicity, in the bifurcation point. Let us assume:

- f is equivariant with respect to the diagonal representation of O(3) on X.
- $f(\lambda, 0) \equiv 0$.
- $(\mathrm{D}f)_{\lambda,0}$ has a pair of complex conjugate eigenvalues $\sigma(\lambda) \pm \mathrm{i}\rho(\lambda)$ with $\sigma(0) = 0$, $\dot{\sigma}(0) \neq 0$, and $\rho(0) = \omega$ of multiplicity $(2l+1) = \mathrm{Dim}(V_l)$ with smooth functions σ and ρ.

The authors now look at subgroups

$$\bar{H} \quad \subset \quad \Gamma.$$

Here the group $\mathrm{S}^1 \subset \Gamma$ acts as a time shift on the periodic solutions. Therefore subgroups $\bar{H}$ consist of spatial and temporal symmetries. For subgroups $\bar{H}$ with

$$\operatorname{Dim}\operatorname{Fix}(\bar{H}) \quad = \quad 2$$

with respect to the representation of the group Γ on $V_l \oplus \mathrm{i}V_l$, the authors prove the existence of exactly one branch of periodic solutions with small amplitude of period near $\frac{2\pi}{\omega}$ and the group of symmetries $\bar{H}$. In order to do this, the authors make a Lyapunov-Schmidt-reduction on the space of periodic functions. The reduced system has the full $\mathrm{O}(3) \times \mathrm{S}^1$-symmetry and solutions correspond to periodic solutions with spatial-temporal symmetries in the original system.
For $l = 2$ [IoRo] applied normal form theory (cf. [EletAl]) to these systems. Up to every finite order they got $\mathrm{O}(3) \times \mathrm{S}^1$-equivariant systems of the form described above. This additional S^1-symmetry up to every finite order is due to the fact that the normal form of f commutes with the one-parameter group $e^{(\mathrm{D}f)^{\mathrm{T}}_{0,0}\, t}$. Due to our conditions on the eigenvalues, this is just a complex rotation.
The following calculations are done using the normal form up to fifth order due to [IoRo]. The normal form up to fifth order is very lengthy and shall not be given here. The parts important for our calculations shall be cited when necessary.
Let G be a compact Lie-group acting on a space X. The most general form of a G-equivariant polynomial mapping $g : X \to X$ is

$$g(x) \quad = \quad \sum_{i=1}^{n} p_i(x)\, e_i(x).$$

Here

$$p_i \quad : \quad X \quad \to \quad \mathbb{R}$$

denote G-invariant polynomials and

$$e_i \quad : \quad X \quad \to \quad X$$

G-equivariant, polynomial mappings.
In order to determine the most general G-equivariant, polynomial mapping up to a fixed order, one, first of all, has to know the number of independend invariants and equivariants and possible relations between them. On this occasion the Poincaré-series described in the next chapter are useful. The next problem is to find the polynomials. In the case of the group $\mathrm{O}(3)$, using raising and lowering operators (cf. [Sa],[Mi]), one can check whether a specific polynomial is invariant or not. The raising and lowering operators are in close relationship to the infinitesimal generators of the Lie-algebra of the group. So the problem is to construct and check all possible polynomials resp. polynomial mappings. Dealing with high order polynomials and large dimensions of the problem, this is a very difficult task that is only accessible via symbolic algebra. At least, using the Poincaré-series, one knows when everything is found.

The lattice of isotropy subgroups of the representation of the group Γ on $V_2 \oplus \mathrm{i}V_2$ has been determined by [MoRoSt].

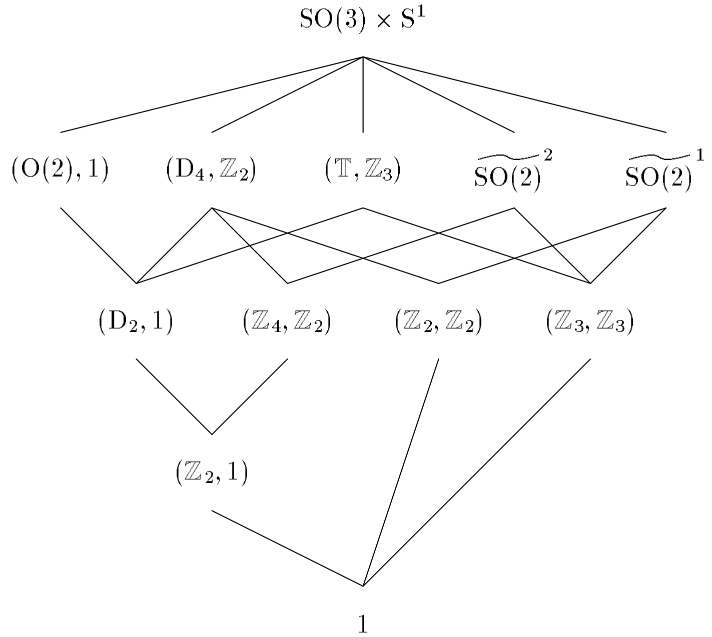


Figure 1: Lattice of isotropy subgroups of Γ on $V_2 \oplus \mathrm{i}V_2$.

The subgroups $\bar{H} \subset \Gamma$ are given as twisted subgroups

$$\bar{H} \quad = \quad \big(H, \Theta(H)\big)$$

with $H \subset \mathrm{SO}(3)$ and $\Theta(H) \subset \mathrm{S}^1$. In this connection

$$\Theta \quad : \quad H \quad \to \quad \mathrm{S}^1$$

is a group homomorphism. Every isotropy subgroup $\bar{H} \subsetneq \Gamma$ can be written in this form (cf. [GoStSch]). In the case of the isotropy subgroups $\widetilde{\mathrm{SO}(2)}^1$ resp. $\widetilde{\mathrm{SO}(2)}^2$ we have $H = \mathrm{SO}(2) \subset \mathrm{SO}(3)$ and $\Theta(H) = \mathrm{S}^1$ with $\Theta(\phi) = \phi$ resp. $\Theta(\phi) = \phi^2$.
In [MoRoSt] the authors investigate Hamiltonian systems of the form

$$\dot{v} \quad = \quad J\,\mathrm{D}H(v)$$

with $v \in \mathbb{R}^{10} = V_2 \oplus \mathrm{i}V_2$,

$$J \quad = \quad \begin{pmatrix} 0 & -I_5 \\ I_5 & 0 \end{pmatrix},$$

and $\mathrm{O}(3) \times \mathrm{S}^1$ invariant Hamiltonian $H : \mathbb{R}^{10} \to \mathbb{R}$. This leads to restrictions on the coefficients of the normal form of the vector field. Like [IoRo] for the general vector field, [MoRoSt] analytically prove the existence of periodic solutions of isotropy

$(\mathrm{O}(2),1)$, $(\mathrm{D}_4,\mathbb{Z}_2)$, $(\mathbb{T},\mathbb{Z}_3)$, $\widetilde{\mathrm{SO}(2)}^1$, and $\widetilde{\mathrm{SO}(2)}^2$. These are exactly the subgroups of Γ having a two dimensional fixed point space for our representation, i.e. the subgroups for which [GoSt] and [GoStSch] predicted the bifurcation of periodic solutions using group theoretical methods. Moreover the authors give conditions for the stability of the different branches of periodic solutions by means of regions in the parameter space of the normal form.

In the following we shall look only at the situation where all solutions bifurcate supercritically. In this case there is a region in parameter space where the periodic solutions of isotropy $(\mathrm{O}(2),1)$ resp. $\widetilde{\mathrm{SO}(2)}^2$ are stable simultaneously, see [IoRo]. Using topological methods, [Le] showed that in this region in parameter space either other isolated invariant objects besides the trivial solution and the different group orbits of periodic solutions have to exist or there is recurrent structure between the trivial solution and the different group orbits of periodic solutions. Recurrent structure means that it is possible to go back via connecting orbits that connect different group orbits in the direction of the flow, from a specific group orbit to this group orbit itself.

In this paper we shall prove the existence of quasiperiodic solutions in the region in parameter space in question. The quasiperiodic solutions given by [IoRo] using the third order normal form cannot be confirmed. We shall prove that the quasiperiodic solutions bifurcate in fifth order from a curve of periodic solutions that is degenerate up to third order.

In order to reduce the dimension of the problem, we shall restrict our calculations in the following to the smallest invariant subspace containing the two stable solutions. This is a subspace of isotropy $(\mathbb{Z}_2,1)$ due to the lattice of isotropy subgroups. Next we want to fix a specific subgroup

$$\mathrm{O}(2) \quad\subset\quad \mathrm{SO}(3)$$

because it is well suited for our coordinates:

$$\mathrm{O}(2) \quad=\quad \left\{ r_\phi = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix},\ \kappa = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix};\ \phi\in[0,2\pi) \right\}.$$

It acts (cf. [GoStSch]) in the following form on our coordinates z:

$$\begin{aligned} r_\phi\,(z_{-2},z_{-1},z_0,z_1,z_2) &= (e^{-2i\phi}z_{-2},e^{-i\phi}z_{-1},z_0,e^{i\phi}z_1,e^{2i\phi}z_2), \\ \kappa\,(z_{-2},z_{-1},z_0,z_1,z_2) &= (z_2,-z_1,z_0,-z_{-1},z_{-2}). \end{aligned}$$

Finally let

$$\Sigma \quad=\quad (\mathbb{Z}_2,1)$$

with

$$\mathbb{Z}_2 \quad=\quad \{1,r_\pi\}.$$

3 Restriction to $\mathrm{Fix}(\mathbb{Z}_2, 1)$

Lemma 3.0.1

$$\mathrm{Fix}(\Sigma) = \mathrm{Span}\{(z_{-2}, 0, z_0, 0, z_2)\} \cong \mathbb{C}^3.$$

Lemma 3.0.2

$$\Xi = \frac{\mathrm{N}(\Sigma)}{\Sigma} = \mathrm{O}(2) \times \mathrm{S}^1.$$

The group $\mathrm{O}(2) \times \mathrm{S}^1$ *acts on* $\mathbb{C}^3$*:*

$$\begin{aligned} r_\theta(z_{-2}, z_0, z_2) &= (e^{-\mathrm{i}\theta} z_{-2}, z_0, e^{\mathrm{i}\theta} z_2), \\ \kappa(z_{-2}, z_0, z_2) &= (z_2, z_0, z_{-2}), \\ \phi(z_{-2}, z_0, z_2) &= (e^{\mathrm{i}\phi} z_{-2}, e^{\mathrm{i}\phi} z_0, e^{\mathrm{i}\phi} z_2). \end{aligned}$$

The group $\mathrm{O}(2)$ *is generated by the rotations* r_θ *and the reflection* κ *and the group* S^1 *by the rotations* ϕ*.*

Proof: We have $\mathrm{N}_{\mathrm{SO}(3)}(\mathbb{Z}_2) = \mathrm{O}(2)$. The representation of $\mathrm{O}(2) \times \mathrm{S}^1$ on $\mathbb{C}^3$ is given by restriction of the representation of $\mathrm{SO}(3) \times \mathrm{S}^1$ on $\mathrm{Fix}(\Sigma)$. ⋈

Let $z = (z_{-2}, z_0, z_2) \in \mathbb{C}^3$. The definition

$$\sigma \overline{z} = \overline{\sigma z}, \ \sigma \in \Xi,$$

gives rise to an unitary representation of Ξ on the space

$$\mathbb{C}^3 \oplus \mathbb{C}^3 \supset \{(z, \overline{z}), \ z \in \mathbb{C}^3\} = \mathbb{R}^6.$$

3.1 Poincare-series, invariants, and equivariants

The number of generators of the ring of Ξ-invariant polynomials $P : \mathbb{R}^6 \to \mathbb{R}$ and of the module of Ξ-equivariant, polynomial mappings $Q : \mathbb{R}^6 \to \mathbb{R}^6$ over the ring of invariant polynomials can be determined using Poincaré-series.

For an unitary representation T of a compact Lie-group G on a vector space V we have

$$\begin{aligned} P_I(t) &= \int_G \frac{1}{\det(I - tT(g))} \mathrm{d}g = \sum_{i=0}^{\infty} c_i t^i, \\ P_{Eq}(t) &= \int_G \frac{\overline{\chi(g)}}{\det(I - tT(g))} \mathrm{d}g = \sum_{i=0}^{\infty} d_i t^i. \end{aligned}$$

Here $c_i, \ i > 0$, denotes the dimension of the vector space of homogeneous invariant polynomials of degree i and $d_i, \ i > 0$, the dimension of the vector space of homogeneous, equivariant mappings of degree i. Let $c_0 = d_0 = 1$. The integral appearing in the formulas is the Haar-integral associated to the compact Lie-group G (cf. [BrtD]), $\chi(g), \ g \in G$, denotes the character of g relative to the representation T. The theory of Poincaré-series is presented in [La2].

LEMMA 3.1.1

$$\begin{aligned} P_I(t) &= \frac{1+t^4}{(1-t^2)^2(1-t^4)^2}, \\ P_{Eq}(t) &= \frac{2t+3t^3+t^5}{(1-t^2)^2(1-t^4)^2}. \end{aligned}$$

Proof: The group $\Xi = \mathrm{O}(2)\times \mathrm{S}^1$ can be written as the disjoint union of two sets in the following form

$$\mathrm{O}(2)\times\mathrm{S}^1 = \mathrm{SO}(2)\times\mathrm{S}^1 \,\dot\cup\, \kappa\mathrm{SO}(2)\times\mathrm{S}^1.$$

Therefore the integrals appearing in the formulas split in two parts.
A. $\Xi_1 = \mathrm{SO}(2)\times\mathrm{S}^1$ acts on the space $\mathbb{C}^3\oplus\mathbb{C}^3$. So we get

$$\begin{aligned} P_I^1(t) &= \int_{\Xi_1} \frac{1}{\det(I-tT(g))}\mathrm{d}g \\ &= \frac{1}{(2\pi)^2}\int_{\phi=0}^{2\pi}\int_{\theta=0}^{2\pi}\frac{1}{\det(I-tT(\theta,\phi))}\mathrm{d}\theta\,\mathrm{d}\phi. \end{aligned}$$

For our representation we have

$$\begin{aligned} \det(I-tT(\theta,\phi)) &= (1-te^{\mathrm{i}(\theta-\phi)})(1-te^{-\mathrm{i}\phi})(1-te^{-\mathrm{i}(\theta+\phi)})(1-te^{\mathrm{i}(-\theta+\phi)}) \\ &\quad (1-te^{\mathrm{i}\phi})(1-te^{\mathrm{i}(\theta+\phi)}). \end{aligned}$$

A transformation of variables

$$e^{\mathrm{i}\theta}\to y_1,\ e^{\mathrm{i}\phi}\to y_2$$

leads to

$$\begin{aligned} P_I^1(t) &= \frac{1}{(2\pi\mathrm{i})^2}\oint_{y_1}\oint_{y_2}\frac{1}{y_1y_2\det(I-tT(y_1,y_2))}\mathrm{d}y_1\,\mathrm{d}y_2 \\ &= \frac{1}{(2\pi\mathrm{i})^2}\oint_{y_1}\oint_{y_2}\frac{y_1y_2^2}{(y_2-ty_1)(y_2-t)(y_1y_2-t)(y_1-ty_2)(1-ty_2)(1-ty_1y_2)}\mathrm{d}y_1\mathrm{d}y_2. \end{aligned}$$

Using the residue theorem twice, one gets

$$P_I^1(t) = \frac{1+t^4}{(1-t^2)^3(1-t^4)}.$$

B. For the set $\kappa\mathrm{SO}(2)\times\mathrm{S}^1$ we have

$$\det(I-tT(\kappa,\theta,\phi)) = (1-te^{-\mathrm{i}\phi})^2(1+te^{-\mathrm{i}\phi})(1-te^{\mathrm{i}\phi})^2(1+te^{\mathrm{i}\phi}).$$

A transformation of variables gives

$$P_I^2(t) = \frac{1}{2\pi\mathrm{i}}\oint_{y_2}\frac{y_2^2}{(y_2-t)^2(y_2+t)(1-ty_2)^2(1+ty_2)}\mathrm{d}y_2.$$

Using the residue theorem, one gets

$$P_I^2(t) \quad = \quad \frac{1+t^4}{(1-t^4)^2(1-t^2)}.$$

C. Because of the normalization of the Haar-integral, we have

$$\begin{aligned} P_I(t) &= \tfrac{1}{2}\left(P_I^1(t)+P_I^2(t)\right) \\ &= \frac{1+t^4}{(1-t^2)^2(1-t^4)^2} \end{aligned}$$

proving the first formula.
D. We want to calculate

$$P_{Eq}^1(t) \quad = \quad \int_{\Xi_1} \frac{\overline{\chi(g)}}{\det(I-tT(g))} \mathrm{d}g.$$

Here we get

$$\begin{aligned} \chi(\theta,\phi) &= \mathrm{Tr}(T(\theta,\phi)) \\ &= e^{\mathrm{i}(-\theta+\phi)}+e^{\mathrm{i}\phi}+e^{\mathrm{i}(\theta+\phi)}+e^{\mathrm{i}(\theta-\phi)}+e^{-\mathrm{i}\phi}+e^{-\mathrm{i}(\theta+\phi)} \\ &= \left(e^{\mathrm{i}\phi}+e^{-\mathrm{i}\phi}\right)\left(e^{\mathrm{i}\theta}+1+e^{-\mathrm{i}\theta}\right). \end{aligned}$$

This leads to

$$\begin{aligned} P_{Eq}^1(t) &= \frac{1}{(2\pi\mathrm{i})^2} \\ &\quad \oint_{y_1}\oint_{y_2} \frac{y_2(1+y_1+y_1^2)(1+y_2^2)}{(y_2-ty_1)(y_2-t)(y_1y_2-t)(y_1-ty_2)(1-ty_2)(1-ty_1y_2)}\mathrm{d}y_1\,\mathrm{d}y_2 \\ &= 2\frac{3t+3t^3}{(1-t^2)^3(1-t^4)}. \end{aligned}$$

E. For the set $\kappa\mathrm{SO}(2)\times\mathrm{S}^1$ one correspondingly gets

$$\chi(\kappa,\theta,\phi) \quad = \quad e^{\mathrm{i}\phi}+e^{-\mathrm{i}\phi}.$$

This leads to

$$\begin{aligned} P_{Eq}^2(t) &= \frac{1}{2\pi\mathrm{i}}\oint_{y_2} \frac{y_2(1+y_2^2)}{(y_2-t)^2(y_2+t)(1-ty_2)^2(1+ty_2)}\mathrm{d}y_2 \\ &= 2\frac{t}{(1-t^2)^2(1-t^4)}. \end{aligned}$$

F. We therefore have

$$\begin{aligned} \tilde{P}_{Eq}(t) &= \tfrac{1}{2}\left(P_{Eq}^1(t)+P_{Eq}^2(t)\right) \\ &= 2\frac{2t+3t^3+t^5}{(1-t^2)^2(1-t^4)^2}. \end{aligned}$$

Doing this, we used the diagonal representation of Ξ on $\mathbb{C}^3 \oplus \mathbb{C}^3$. But we are interested in the subspace $\{(z,\overline{z}),\ z \in \mathbb{C}^3\} \subset \mathbb{C}^3 \oplus \mathbb{C}^3$ only. Therefore the number of equivariants given by the formula is twice as big as it should be counting also equivariants with one component being zero. $\bowtie$

The Poincaré-series can be interpreted in the following way.

Lemma 3.1.2 *The polynomials*

$$\begin{aligned}
\pi_1 &= |z_0|^2, \\
\pi_2 &= |z_{-2}|^2 + |z_2|^2, \\
\pi_3 &= |z_{-2}|^2 |z_2|^2, \\
\pi_4 &= \frac{1}{2}\left(\overline{z_0}^2 z_{-2} z_2 + z_0^2 \overline{z_{-2}}\,\overline{z_2}\right), \\
\pi_5 &= \frac{\mathrm{i}}{2}\left(\overline{z_0}^2 z_{-2} z_2 - z_0^2 \overline{z_{-2}}\,\overline{z_2}\right)
\end{aligned}$$

are a minimal set of generators of the ring of invariant polynomials.

$$P \quad : \quad \mathbb{R}^6 \quad \to \quad \mathbb{R}.$$

The only relation between them is

$$\pi_4^2 + \pi_5^2 \quad = \quad \pi_1^2 \pi_3.$$

Proof: One easily sees that the given polynomials $\pi_1, \ldots, \pi_5$ are invariant, and just meet the given relation. Therefore the Poincaré-series of these polynomials is identical to the one calculated. Because of this there are no additional generators and relations. $\bowtie$

Introducing polar coordinates in the following form

$$z_j = r_j e^{\mathrm{i}\phi_j},\ j \in \{-2, 0, 2\},$$

and defining

$$\theta \quad = \quad 2\phi_0 - \phi_{-2} - \phi_2,$$

one gets

$$\pi_4 \quad = \quad r_0^2 r_{-2} r_2 \cos\theta$$

and

$$\pi_5 \quad = \quad r_0^2 r_{-2} r_2 \sin\theta.$$

Consequently the invariants π_4 and π_5 represent phase relations between the different coordinates.

Lemma 3.1.3 *Let $\pi : \mathbb{R}^6 \to \mathbb{R}$ be an invariant polynomial for the representation of Ξ on $\mathbb{R}^6$.*
Then

$$p(z,\overline{z}) = \overline{\nabla_{z,\overline{z}} \pi(z,\overline{z})}$$

is a Ξ-equivariant polynomial mapping for this representation.

Proof: We have

$$p(\sigma(z,\overline{z})) = \overline{\nabla_{\sigma(z,\overline{z})}\pi(z,\overline{z})} = \overline{\nabla_{z,\overline{z}}\pi(z,\overline{z})}\,\overline{\sigma^{-1}} = \sigma p(z,\overline{z}).$$

The last equality is correct because the representation is unitary. ⋈

LEMMA 3.1.4 *The independent, Ξ-equivariant, polynomial mappings*

$$Q \quad : \quad \mathbb{R}^6 \; \rightarrow \; \mathbb{R}^6$$

up to fifth order are

$$e_1 = \begin{pmatrix} 0 \\ z_0 \\ 0 \end{pmatrix}, \; e_2 = \begin{pmatrix} z_{-2} \\ 0 \\ z_2 \end{pmatrix}, \; e_3 = \begin{pmatrix} z_{-2}|z_2|^2 \\ 0 \\ z_2|z_{-2}|^2 \end{pmatrix},$$

$$e_4 = \frac{1}{2}\begin{pmatrix} z_0^2\overline{z_2} \\ 2z_{-2}z_2\overline{z_0} \\ z_0^2\overline{z_{-2}} \end{pmatrix}, \; e_5 = -\frac{\mathrm{i}}{2}\begin{pmatrix} z_0^2\overline{z_2} \\ -2z_{-2}z_2\overline{z_0} \\ z_0^2\overline{z_{-2}} \end{pmatrix}.$$

Here e_i, $i = 1,\ldots,5$, always denote the first component of the equivariant. The second is given by complex conjugation of the first one.

Proof: Using the previous lemma, one knows that the mappings $e_j = \overline{\nabla_{z,\overline{z}}\pi_j}$, $j = 1,\ldots,5$, are equivariant. Power series expansion of $P_{Eq}(t)$ leads to

$$P_{Eq}(t) = 2t + 7t^3 + 17t^5 + \mathrm{O}(t^7).$$

There are $2, 7$ resp. 18 different possibilities to construct equivariant mappings of degree $1, 3$ resp. 5 from invariant polynomials $\pi_1,\ldots,\pi_5$ and equivariant mappings $e_1,\ldots,e_5$ by multiplication of invariants with an equivariant. In the fifth order one gets the relation

$$e_1(\pi_4 - \mathrm{i}\pi_5) \quad = \quad \frac{1}{2}\pi_1(e_4 - \mathrm{i}e_5).$$

All other combinations can't be generated this way. Therefore the Poincaré-series belonging to $\pi_1,\ldots,\pi_5$ and $e_1,\ldots,e_5$ is identical to the calculated one up to fifth order. Because of this there are no further generators or relations up to fifth order. ⋈

3.2 ORBIT SPACE REDUCTION

The most general $\mathrm{O}(2)\times\mathrm{S}^1$-equivariant Hopf-bifurcation problem on $\mathbb{R}^6$ up to third order has the form

$$\dot{z} \quad = \quad (\lambda + \mathrm{i}\omega)(e_1 + e_2) + a_1\pi_1 e_1 + a_2\pi_1 e_2 + a_3\pi_2 e_1 + a_4\pi_2 e_2 + a_5 e_3 + a_6 e_4 + a_7 e_5,$$

$a_j \in \mathbb{C}$, $j = 1,\ldots,7$, $\lambda,\omega \in \mathbb{R}$, and $z = (z_{-2}, z_0, z_2)$.
We want to study bifurcation problems on $\mathbb{R}^6$ resulting from a $\mathrm{SO}(3)\times\mathrm{S}^1$-equivariant

problem on $V_2 \oplus iV_2$. This gives the following restrictions for the coefficients $a_1, \ldots, a_7$:

$$\begin{aligned} \dot{z} &= (\lambda + i\omega)(e_1 + e_2) + \left(a - \frac{1}{2}b - \sqrt{\frac{3}{2}}c\right)\pi_1 e_1 + \left(a - \sqrt{\frac{8}{3}}c\right)\pi_1 e_2 \\ &\quad + \left(a - \sqrt{\frac{8}{3}}c\right)\pi_2 e_1 + a\pi_2 e_2 - \left(b + \sqrt{6}c\right)e_3 + \left(-b + \sqrt{\frac{2}{3}}c\right)e_4 \\ &\quad + 0e_5. \end{aligned} \tag{3.2.1}$$

Here $a, b, c \in \mathbb{C}$ denote the corresponding coefficients from the normal form of [IoRo]. This is obtained by comparison of the normal form of [IoRo] restricted to the subspace with the general equation. Define coefficients $\alpha, \beta, \gamma \in \mathbb{C}$:

$$\begin{aligned} \alpha &= a - \tfrac{1}{2}b - \sqrt{\tfrac{3}{2}}c, & a &= \gamma, \\ \beta &= a - \sqrt{\tfrac{8}{3}}c, & b &= -2\alpha + \tfrac{3}{2}\beta + \tfrac{1}{2}\gamma, \\ \gamma &= a, & c &= \sqrt{\tfrac{3}{8}}(\gamma - \beta). \end{aligned}$$

Then the vector field has the form

$$\begin{aligned} \dot{z} &= (\lambda + i\omega)(e_1 + e_2) + \alpha\pi_1 e_1 + \beta(\pi_1 e_2 + \pi_2 e_1) + \gamma\pi_2 e_2 \\ &\quad + 2(\alpha - \gamma)e_3 + 2(\alpha - \beta)e_4 \\ &= \Big((\lambda + i\omega) + \alpha\pi_1 + \beta\pi_2\Big)e_1 + \Big((\lambda + i\omega) + \beta\pi_1 + \gamma\pi_2\Big)e_2 \\ &\quad + 2(\alpha - \gamma)e_3 + 2(\alpha - \beta)e_4 \end{aligned} \tag{3.2.2}$$

with $\lambda, \omega \in \mathbb{R}$.

Let $\dot{x} = f(x)$ be a differential equation on a vector space X. Let the mapping f be equivariant with respect to the representation of the compact Lie-group G on X. Since

$$\dot{(gx)} = g\dot{x} = gf(x) = f(gx), \ \forall g \in G,$$

$gx(t),\ g \in G$, is a solution if $x(t)$ is a solution. This means one has to deal with group orbits Gx of solutions. Let G_x denote the isotropy of a point x. Then we have

$$\frac{G}{G_x} \cong Gx.$$

Here $\frac{G}{G_x}$ and Gx are compact manifolds and we have (cf. [Di])

$$\operatorname{Dim} Gx = \operatorname{Dim} G - \operatorname{Dim} G_x.$$

In order to get rid of the redundancy in our system induced by the group G, one studies the orbit space $\frac{X}{G}$. Here points lying on a group orbit are identified:

$$x \simeq y \iff x = gy \text{ with } x, y \in X \text{ and } g \in G.$$

The orbit space is homeomorphic to the image of the Hilbert-map $\Pi(X)$

$$\begin{aligned} \Pi \ : \ X &\to \mathbb{R}^k \\ x &\to (\pi_i(x)) \end{aligned}$$

(cf. [La2], [Bi]). Here k denotes the minimal number of generators of the ring of G-invariant polynomials $P : X \to \mathbb{R}$ and π_i, $i = 1, \ldots, k$, is such a system of generators. The original differential equation is reduced to a differential equation on $\Pi(X)$ of the form

$$\dot{\pi} = g(\pi) \text{ with } \pi = (\pi_1, \ldots, \pi_k).$$

The reduced equation can be calculated as follows:

$$\dot{\pi_i} = \nabla_x \pi_i \, \dot{x} = \nabla_x \pi_i \, f(x), \;\; i = 1, \ldots, k.$$

The advantage of this reduction lies in the fact that in general the dimension of the reduced problem is smaller than the original one. Furthermore symmetry induced periodic solutions in the original system correspond to fixed points in the reduced system and can be dealt with more easily analytically. The disadvantage is that the orbit space in general is no vector space but a stratified space.

In our case the differential equation up to third order (Equation (3.2.2)) is given in the form

$$\dot{z} \;\; = \;\; \sum_{j=1}^{5} q_j \, e_j .$$

Here

$$q_j : \mathbb{R}^6 \to \mathbb{C}, \;\; j = 1, \ldots, 5,$$

are invariant polynomials. So one gets

$$\begin{aligned} \dot{\pi_i} &= \nabla_z \pi_i \, \dot{z} + \nabla_{\overline{z}} \pi_i \, \dot{\overline{z}} \\ &= \overline{e_i} \, \dot{z} + e_i \, \dot{\overline{z}} \\ &= 2\mathrm{Re}\, (\overline{e_i} \, \dot{z}) \\ &= 2\mathrm{Re} \left(\sum_{j=1}^{5} q_j \, \overline{e_i} e_j \right) . \end{aligned}$$

The products $\overline{e_i} e_j$, $i \le j \in \{1, \ldots, 5\}$, are

$$\begin{array}{ll} \overline{e_1} e_1 = \pi_1 & \overline{e_2} e_2 = \pi_2 \\ \overline{e_1} e_2 = 0 & \overline{e_2} e_3 = 2\pi_3 \\ \overline{e_1} e_3 = 0 & \overline{e_2} e_4 = \pi_4 + \mathrm{i}\pi_5 \\ \overline{e_1} e_4 = \pi_4 - \mathrm{i}\pi_5 & \overline{e_2} e_5 = -\mathrm{i}\pi_4 + \pi_5 \\ \overline{e_1} e_5 = \mathrm{i}\pi_4 + \pi_5 & \end{array}$$

$$\begin{array}{lll} \overline{e_3} e_3 = \pi_2 \pi_3 & \overline{e_4} e_4 = \frac{1}{4}\pi_1^2 \pi_2 + \pi_1 \pi_3 & \overline{e_5} e_5 = \frac{1}{4}\pi_1^2 \pi_2 + \pi_1 \pi_3 . \\ \overline{e_3} e_4 = \frac{1}{2}\pi_2 (\pi_4 + \mathrm{i}\pi_5) & \overline{e_4} e_5 = -\frac{\mathrm{i}}{4}\pi_1^2 \pi_2 + \mathrm{i}\pi_1 \pi_3 & \\ \overline{e_3} e_5 = \frac{1}{2}\pi_2 (-\mathrm{i}\pi_4 + \pi_5). & & \end{array}$$

For $i > j \in \{1, \ldots, 5\}$ we have

$$\overline{e_i} e_j \;\; = \;\; \overline{\overline{e_j} e_i} .$$

So the following lemma is proved.

Lemma 3.2.1 *The Vector Field (3.2.2) yields the following reduced vector field on the orbit space*

$$\begin{aligned}
\dot{\pi_1} &= 2(\lambda+\alpha^r\pi_1+\beta^r\pi_2)\pi_1+4\Big((\alpha-\beta)^r\pi_4+(\alpha-\beta)^i\pi_5\Big)\\
\dot{\pi_2} &= 2(\lambda+\beta^r\pi_1+\gamma^r\pi_2)\pi_2+8(\alpha-\gamma)^r\pi_3+4\Big((\alpha-\beta)^r\pi_4-(\alpha-\beta)^i\pi_5\Big)\\
\dot{\pi_3} &= 4(\lambda+\beta^r\pi_1+\alpha^r\pi_2)\pi_3+2\pi_2\Big((\alpha-\beta)^r\pi_4-(\alpha-\beta)^i\pi_5\Big)\\
\dot{\pi_4} &= 2\Big(2\lambda+(\alpha+\beta)^r(\pi_1+\pi_2)\Big)\pi_4+2(\alpha-\beta)^i(-\pi_1+\pi_2)\pi_5\\
&\quad +(\alpha-\beta)^r\pi_1(\pi_1\pi_2+4\pi_3)\\
\dot{\pi_5} &= 2\Big(2\lambda+(\alpha+\beta)^r(\pi_1+\pi_2)\Big)\pi_5+2(\alpha-\beta)^i(\pi_1-\pi_2)\pi_4\\
&\quad +(\alpha-\beta)^i\pi_1(-\pi_1\pi_2+4\pi_3).
\end{aligned}$$

Here $\alpha^r,\beta^r,\gamma^r$ *resp.* $\alpha^i,\beta^i,\gamma^i$ *denote the real resp. imaginary parts of* α,β,γ.

3.3 Lattice of isotropy subgroups

All isotropy subgroups $G \subsetneq \mathrm{O}(2)\times \mathrm{S}^1$ can be written as twisted subgroups in the form

$$G = H^{\Theta} = \{(h,\Theta(h)) \in \mathrm{O}(2)\times\mathrm{S}^1 \mid h\in H\}$$

(cf. [GoSt], [GoStSch]). Here $H\subset \mathrm{O}(2)$ denotes a closed subgroup of $\mathrm{O}(2)$ and

$$\Theta \quad : \quad \mathrm{O}(2) \quad \to \quad \mathrm{S}^1$$

is a group homomorphism. For a closed subgroup $H\subset\mathrm{O}(2)$ let

$$H' = \big\langle g^{-1}h^{-1}gh \mid g,h\in H\big\rangle$$

denote the commutator of H and

$$H^{ab} = \frac{H}{H'}$$

the abelianisation of H. Since $\Theta(H)\subset\mathrm{S}^1$ is abelian, the possible twist typs $\Theta(H)$ of H can be concluded from the abelianisation H^{ab}. One gets the following table.

H	H'	H^{ab}	$\Theta(H)$
$\mathrm{O}(2)$	$\mathrm{SO}(2)$	$\mathbb{Z}_2$	$1,\mathbb{Z}_2$
$\mathrm{SO}(2)$	1	$\mathrm{SO}(2)$	$1,\mathrm{S}^1$
$\mathrm{D_n}$	$\mathbb{Z}_{\frac{n}{2}}$, n even $\mathbb{Z}_{\mathrm{n}}$, n odd	$\mathbb{Z}_2\oplus\mathbb{Z}_2$, n even $\mathbb{Z}_2$, n odd	$1,\mathbb{Z}_2$
$\mathbb{Z}_{\mathrm{n}}$	1	$\mathbb{Z}_{\mathrm{n}}$	$1,\mathbb{Z}_{\mathrm{d}}$, $d\|n$

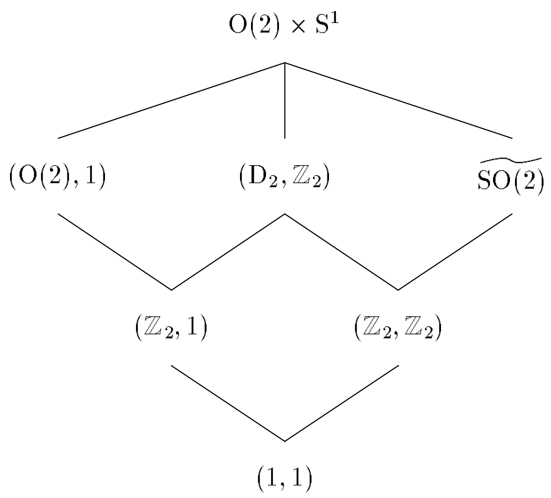


Figure 2: Lattice of isotropy subgroups of $\mathrm{O}(2) \times \mathrm{S}^1$ on $\mathbb{R}^6$

Lemma 3.3.1 *For our representation of the group* $\mathrm{O}(2) \times \mathrm{S}^1$ *on the space* $\mathbb{R}^6$ *one gets the following lattice of isotropy subgroups.*
The following table contains generating elements, representatives and the dimension of the associated fixed point space for every group H^{Θ}.

H^{Θ}	generators	representative	$\mathrm{Dim\,Fix}(H^{\Theta})$
$\mathrm{O}(2) \times \mathrm{S}^1$	$\mathrm{O}(2) \times \mathrm{S}^1$	$(0, 0, 0)$	0
$(\mathrm{O}(2), 1)$	$(\mathrm{O}(2), 1)$	$(0, z_0, 0)$	2
$\widetilde{\mathrm{SO}(2)}$	$\langle (\phi, \phi),\ \phi \in \mathrm{S}^1 \rangle$	$(z_{-2}, 0, 0)$	2
$(\mathrm{D}_2, \mathbb{Z}_2)$	$\langle (\kappa, 1), (\pi, \pi) \rangle$	$(z_2, 0, z_2)$	2
$(\mathbb{Z}_2, \mathbb{Z}_2)$	$\langle (\pi, \pi) \rangle$	$(z_{-2}, 0, z_2)$	4
$(\mathbb{Z}_2, 1)$	$\langle (\kappa, 1) \rangle$	(z_2, z_0, z_2)	4
$(1, 1)$	$\{(1, 1)\}$	(z_{-2}, z_0, z_2)	6

Proof: The dimension of the fixed point space of a potential isotropy subgroup

$$H^{\Theta} \quad \subset \quad \mathrm{O}(2) \times \mathrm{S}^1$$

is given by the trace formula (cf. [GoSt], [GoStSch])

$$\mathrm{Dim\,Fix}\, H^{\Theta} \quad = \quad \int_{H^{\Theta}} \mathrm{Tr}(h, \theta(h))\, \mathrm{d}h.$$

The values of $\mathrm{Tr}(h,\theta(h))$, $h \in \mathrm{O}(2)$, $\Theta(h) \in \mathrm{S}^1$, are known by Section 3.1. Since we use the diagonal representation of the group $\mathrm{O}(2) \times \mathrm{S}^1$ on $\mathbb{C}^3 \oplus \mathbb{C}^3 \supset \mathbb{R}^6$ the formula yields the real dimension of the fixed point space.

A. Let $\Theta(H) = 1$. Then

$$\begin{aligned}
\operatorname{Dim}\operatorname{Fix}(\mathrm{O}(2),1) &= \frac{1}{2}\left(\frac{1}{2\pi}\int_{\delta=0}^{2\pi} 2(1+2\cos\delta)\,\mathrm{d}\delta + \int_{\delta=0}^{2\pi} 2\,\mathrm{d}\delta\right) = 2,\\
\operatorname{Dim}\operatorname{Fix}(\mathrm{SO}(2),1) &= \frac{1}{2\pi}\int_{\phi=0}^{2\pi} 2(1+2\cos\phi)\,\mathrm{d}\phi = 2,\\
\operatorname{Dim}\operatorname{Fix}(\mathrm{D_n},1) &= \frac{1}{2n}\left(\sum_{j=1}^{n} 2\left(1+2\cos\frac{2\pi}{n}j\right) + \sum_{j=1}^{n} 2\right) = \begin{cases} 4 & n=1,\\ 2 & n\geq 2,\end{cases}\\
\operatorname{Dim}\operatorname{Fix}(\mathbb{Z}_\mathrm{n},1) &= \frac{1}{n}\sum_{j=1}^{n} 2\left(1+2\cos\frac{2\pi}{n}j\right) = 2.
\end{aligned}$$

The subspaces $\{(0,z_0,0)\}$ resp. $\{(z_2,z_0,z_2)\}$ have isotropy $(\mathrm{O}(2),1)$ resp. $(\mathbb{Z}_2,1)$ and, consequently, $(\mathrm{O}(2),1)$ resp. $(\mathbb{Z}_2,1)$ are isotropy subgroups with two resp. four dimensional fixed point spaces. Let $\mathbb{Z}_2 = \mathrm{D}_1$ denote the $\mathbb{Z}_2$ generated by κ. The other groups with trivial twist are no isotropy subgroups.

B. Let $\Theta(H) = \mathrm{S}^1$. Possible twists are

$$\begin{array}{rcccl} \Theta_k & : & \mathrm{SO}(2) & \rightarrow & \mathrm{S}^1\\ & & \phi & \rightarrow & k\phi \end{array}$$

with $k \in \mathbb{N}$. Then we have

$$\operatorname{Dim}\operatorname{Fix}\widetilde{\mathrm{SO}(2)}^{k} = \frac{1}{2\pi}\int_{\phi=0}^{2\pi} 2(1+2\cos\phi)\cos k\phi\,\mathrm{d}\phi = \begin{cases} 2 & k=1,\\ 0 & k>1.\end{cases}$$

The subspace $\{(z_{-2},0,0)\}$ has isotropy $\widetilde{\mathrm{SO}(2)}$ and, therefore, $\widetilde{\mathrm{SO}(2)}$ is an isotropy group with two dimensional fixed point space.

C. Let $\Theta(H) = \mathbb{Z}_2$. Then

$$\operatorname{Dim}\operatorname{Fix}(\mathrm{O}(2),\mathbb{Z}_2) = \frac{1}{2}\left(\frac{1}{2\pi}\int_{\delta=0}^{2\pi} 2(1+2\cos\delta)\,\mathrm{d}\delta - \int_{\delta=0}^{2\pi} 2\,\mathrm{d}\delta\right) = 0.$$

In the case $(\mathrm{D_n},\mathbb{Z}_2)$ there are several possibilities. Let first n be even. Here we have three possible twists.

To begin with let

$$H^{\Theta_{1,n}} = \left\langle \left(\frac{2\pi}{n},\pi\right), (\kappa,1)\right\rangle.$$

Then

$$\begin{aligned}
\operatorname{Dim}\operatorname{Fix} H^{\Theta_{1,n}} &= \frac{1}{2n}\left(\sum_{j=1}^{n} 2(-1)^j\left(1+2\cos\frac{2\pi}{n}j\right) + \sum_{j=1}^{n} 2(-1)^j\right)\\
&= \begin{cases} 2 & n=2,\\ 0 & n\geq 4.\end{cases}
\end{aligned}$$

Defining

$$H^{\Theta_{2,n}} = \left\langle \left(\frac{2\pi}{n}, \pi\right), (\kappa, \pi) \right\rangle,$$

we have

$$\begin{aligned} \operatorname{Dim}\operatorname{Fix} H^{\Theta_{2,n}} &= \frac{1}{2n}\left(\sum_{j=1}^{n} 2(-1)^j\left(1 + 2\cos\frac{2\pi}{n}j\right) + \sum_{j=1}^{n} 2(-1)^{j+1}\right) \\ &= \begin{cases} 2 & n = 2, \\ 0 & n \geq 4. \end{cases} \end{aligned}$$

Finally let

$$H^{\Theta_{3,n}} = \left\langle \left(\frac{2\pi}{n}, 1\right), (\kappa, \pi) \right\rangle.$$

Then

$$\operatorname{Dim}\operatorname{Fix} H^{\Theta_{3,n}} = \frac{1}{2n}\left(\sum_{j=1}^{n} 2\left(1 + 2\cos\frac{2\pi}{n}j\right) + \sum_{j=1}^{n} -2\right) = 0.$$

Setting

$$(\mathrm{D}_2, \mathbb{Z}_2) = \langle(\pi, \pi), (\kappa, 1)\rangle = H^{\Theta_{1,2}},$$

we have

$$\left(-\frac{\pi}{2}, 1\right) H^{\Theta_{2,2}} \left(\frac{\pi}{2}, 1\right) = H^{\Theta_{1,2}}.$$

Therefore both groups are conjugated.
The subspace $\{(z_2, 0, z_2)\}$ has isotropy $(\mathrm{D}_2, \mathbb{Z}_2)$ and, therefore, $(\mathrm{D}_2, \mathbb{Z}_2)$ is an isotropy group with two dimensional fixed point space.
If n is odd, then

$$\operatorname{Dim}\operatorname{Fix}(\mathrm{D_n}, \mathbb{Z}_2) = \frac{1}{2n}\left(\sum_{j=1}^{n} 2\left(1 + 2\cos\frac{2\pi}{n}j\right) + \sum_{j=1}^{n} -2\right) = \begin{cases} 2 & n = 1, \\ 0 & n \geq 3. \end{cases}$$

$(\mathrm{D}_1, \mathbb{Z}_2) = \langle(\kappa, \pi)\rangle$ is extended by $H^{\Theta_{2,2}}$ and, consequently, is no isotropy group.
In the case $(\mathbb{Z}_\mathrm{n}, \mathbb{Z}_2)$, in particular n has to be even, we have

$$\operatorname{Dim}\operatorname{Fix}(\mathbb{Z}_\mathrm{n}, \mathbb{Z}_2) = \frac{1}{n}\sum_{j=1}^{n} 2(-1)^j\left(1 + 2\cos\frac{2\pi}{n}j\right) = \begin{cases} 4 & n = 2, \\ 0 & n \geq 4. \end{cases}$$

The subspace $\{(z_{-2}, 0, z_2)\}$ has isotropy $(\mathbb{Z}_2, \mathbb{Z}_2) = \langle(\pi, \pi)\rangle$ and, therefore, $(\mathbb{Z}_2, \mathbb{Z}_2)$ is an isotropy group with four dimensional fixed point space.
D. Finally we have to study the case $(\mathbb{Z}_\mathrm{n}, \mathbb{Z}_\mathrm{d})$ with $d|n$ and $n \geq 2$. Possible nontrivial twists for $\mathbb{Z}_\mathrm{n}$ are

$$\begin{array}{rccc} \Theta_k : & \mathbb{Z}_\mathrm{n} & \to & \mathrm{S}^1 \\ & \frac{2\pi}{n}j & \to & \frac{2\pi}{n}jk \end{array}$$

with $1 \leq k < n$. This gives

$$\begin{aligned}\operatorname{Dim}\operatorname{Fix}(\mathbb{Z}_\mathrm{n}, \Theta_k(\mathbb{Z}_\mathrm{n})) &= \frac{1}{n}\sum_{j=1}^{n} 2\left(1 + 2\cos\frac{2\pi}{n}j\right)\cos\frac{2\pi}{n}jk \\ &= \begin{cases} 4 & n = 2,\ k = 1 \\ 2 & n \geq 3,\ k \in \{1, n-1\} \\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Studying the representations T_k of $\mathrm{D_n} = \langle\sigma, \kappa\rangle$ on $\mathbb{C}^2$ with

$$T_k(\sigma) = \begin{pmatrix} e^{-\mathrm{i}\frac{2\pi}{n}k} & 0 \\ 0 & e^{\mathrm{i}\frac{2\pi}{n}k}\end{pmatrix}$$

and

$$T_k(\kappa) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

the last equality follows. The representations T_k are irreducible for $n \geq 3$. The representations T_1 and T_{n-1} are conjugated since

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} e^{-\mathrm{i}\frac{2\pi}{n}} & 0 \\ 0 & e^{\mathrm{i}\frac{2\pi}{n}}\end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} e^{\mathrm{i}\frac{2\pi}{n}} & 0 \\ 0 & e^{-\mathrm{i}\frac{2\pi}{n}}\end{pmatrix}.$$

Orthogonality relations for these representations (cf. [La2]) yield the equality.
The case $(\mathbb{Z}_2, \mathbb{Z}_2)$ has been dealt with in part c of the proof, the other cases correspond to conjugated twists of typ

$$\begin{array}{cccc}\Theta_k & : & \mathbb{Z}_\mathrm{n} & \to & \mathrm{S}^1 \\ & & \frac{2\pi}{n}j & \to & \pm\frac{2\pi}{n}j.\end{array}$$

These are extended by the isotropy group $\widetilde{\mathrm{SO}(2)}$. $\bowtie$

Lemma 3.3.2 *For the isotropy groups $H^\Theta \subset \mathrm{SO}(3)\times\mathrm{S}^1$ introduced in the first chapter we have*

H^Θ	$\frac{H^\Theta\cap\mathrm{N}(\Sigma)}{\Sigma}$
$(\mathbb{Z}_2, 1)$	$(1, 1)$
$(\mathbb{Z}_4, \mathbb{Z}_2)$	$(\mathbb{Z}_2, \mathbb{Z}_2)$
$(\mathrm{D}_2, 1)$	$(\mathbb{Z}_2, 1)$
$(\mathrm{O}(2), 1)$	$(\mathrm{O}(2), 1)$
$(\mathrm{D}_4, \mathbb{Z}_2)$	$(\mathrm{D}_2, \mathbb{Z}_2)$
$\widetilde{\mathrm{SO}(2)}^2$	$\widetilde{\mathrm{SO}(2)}$
$(\mathbb{T}, \mathbb{Z}_3)$	$(\mathbb{Z}_2, 1)$.

Note that

$$H^\Theta \subset N(\Sigma)$$

for all isotropy groups H^{Θ} except for the group $(\mathbb{T}, \mathbb{Z}_3)$. The group $(\mathbb{T}, \mathbb{Z}_3)$ does not correspond to a special isotropy typ in the $\mathrm{O}(2) \times \mathrm{S}^1$-equivariant system. But the restricted Vector Field (3.2.2) leaves the corresponding two dimensional fixed point space lying in $\mathrm{Fix}\,(\Sigma)$ invariant.

LEMMA 3.3.3

$$\mathrm{Fix}\,(\mathbb{T}, \mathbb{Z}_3) \;=\; \left\{ \left(\frac{\mathrm{i}}{\sqrt{2}} z_0, z_0, \frac{\mathrm{i}}{\sqrt{2}} z_0 \right),\ z_0 \in \mathbb{C} \right\}.$$

Proof: Using the representation of $\mathrm{SO}(3)$ on the space $V_2 \oplus \mathrm{i} V_2$ introduced in the first chapter, one gets the following representation of the group

$$\mathbb{T} \;=\; \langle \pi, \tau \rangle \;\subset\; \mathrm{SO}(3)$$

with

$$\tau \;=\; \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

on the subspace $\{(z_2, z_0, z_2)\} \subset \mathbb{R}^6$:

$$\begin{aligned} \pi(z_2, z_0, z_2) &= (z_2, z_0, z_2), \\ \tau(z_2, z_0, z_2) &= \left(-\frac{1}{2} z_2 - \frac{1}{2}\sqrt{\frac{3}{2}} z_0, \sqrt{\frac{3}{2}} z_2 - \frac{1}{2} z_0, -\frac{1}{2} z_2 - \frac{1}{2}\sqrt{\frac{3}{2}} z_0 \right). \end{aligned}$$

If an element has the form

$$\left\{ \left(\frac{\mathrm{i}}{\sqrt{2}} z_0, z_0, \frac{\mathrm{i}}{\sqrt{2}} z_0 \right),\ z_0 \in \mathbb{C} \right\},$$

then

$$\left(\tau, e^{\mathrm{i}\frac{2\pi}{3}} \right) (z_2, z_0, z_2) \;=\; (z_2, z_0, z_2).$$

⋈

3.4 CRITICAL POINTS OF THE REDUCED VECTOR FIELD

LEMMA 3.4.1 *The image of the Hilbert-map $\Pi(\mathbb{R}^6)$ is sketched in Figure 3. One has to imagine circles of radius*

$$\pi_4^2 + \pi_5^2 = \pi_1^2 \pi_3$$

attached to points of the sketch. We have the following assignment

$(\pi_1, \ldots, \pi_5) \in \Pi(\mathbb{R}^6)$	*isotropy typ*
π_1-*axis*	$(\mathrm{O}(2), 1)$
π_2-*axis*	$\widetilde{\mathrm{SO}(2)}$
$\pi_1 = 0,\ \pi_3 = \frac{1}{4}\pi_2^2$	$(\mathrm{D}_2, \mathbb{Z}_2)$
$\pi_1 = 0,\ 0 < \pi_3 < \frac{1}{4}\pi_2^2$	$(\mathbb{Z}_2, \mathbb{Z}_2)$
$\pi_1 > 0,\ \pi_3 = \frac{1}{4}\pi_2^2$	$(\mathbb{Z}_2, 1)$
$\pi_1 > 0,\ 0 \leq \pi_3 < \frac{1}{4}\pi_2^2$	$(1, 1)$.

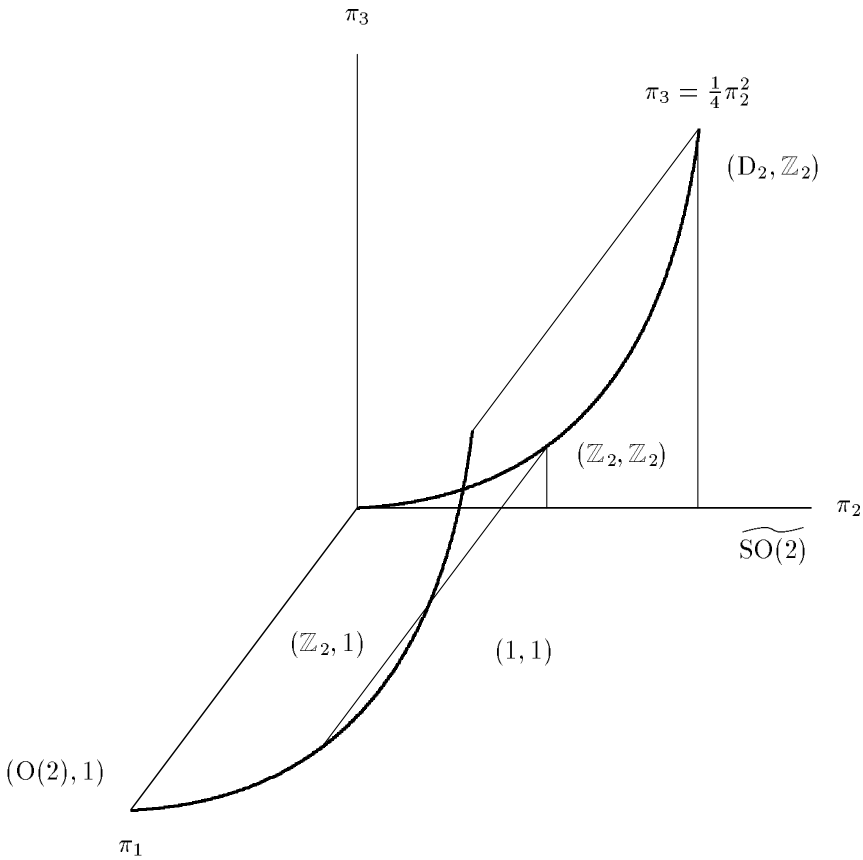


Figure 3: Image of the Hilbert-map

REMARK 3.4.2 *In the following the image of the Hilbert-map* $\Pi(\mathbb{R}^6)$ *shall be denoted Hilbert-set. Since the invariants* π_1, π_2, *and* π_3 *by definition mean radii, only non-negative values are possible. In* (π_1, π_2, π_3)*-space the Hilbert-set is a wedge (cf. Figure 3) limited at the top by the surface* $\pi_3 = \frac{1}{4}\pi_2^2$, *at the bottom by the surface* $\pi_3 = 0$, *and at the back by the surface* $\pi_1 = 0$.

Proof: By definition of the invariants in Lemma 3.1.2 we have

$$\pi_1, \ \pi_2, \ \pi_3 \geq 0.$$

A calculation using Lagrange-multipliers yields the possible values of π_3

$$0 \leq \pi_3 \leq \frac{1}{4}\pi_2^2.$$

The relation

$$\pi_4^2 + \pi_5^2 = \pi_1^2 \pi_3$$

has to be satisfied by Lemma 3.1.2. ⋈

REMARK 3.4.3 *Points with isotropy* $(\mathrm{O}(2), 1)$ *and* $(\mathrm{D}_2, \mathbb{Z}_2)$ *and images of points with isotropy* $(\mathbb{T}, \mathbb{Z}_3)$ *in the original system (cf. Lemma 3.4.5) under the Hilbert-map satisfy the relation*

$$\Delta \;=\; \frac{1}{4}\pi_2^2 - \pi_3 \;=\; 0.$$

In the following we shall study the reduced vector field (cf. Lemma 3.2.1) on the Hilbert-set $\Pi(\mathbb{R}^6)$.

LEMMA 3.4.4 *Let*

$$\Delta \;=\; \frac{1}{4}\pi_2^2 - \pi_3.$$

Then

$$\dot{\Delta} \;=\; 4\,\Delta\,(\lambda + \beta^r\pi_1 + \gamma^r\pi_2)\,.$$

Proof: The stratum

$$\Delta \;=\; 0$$

corresponds to points with a certain isotropy and, therefore, is flow invariant. Thus we have $\dot{\Delta} = 0$ for $\Delta = 0$ and there exists a relation of the form

$$\dot{\Delta} \;=\; \Delta\, r(\pi_1, \ldots, \pi_5).$$

A simple calculation gives the precise relation. ⋈

LEMMA 3.4.5 *The orbit space reduction maps* Fix $(\mathbb{T}, \mathbb{Z}_3)$ *to the invariant curve*

$$\left(\pi_1, \pi_1, \frac{1}{4}\pi_1^2, -\frac{1}{2}\pi_1^2, 0\right) \subset \Pi(\mathbb{R}^6),\;\; \pi_1 > 0,$$

located on the stratum $\Delta = 0$.

Proof: The proof follows directly from the Lemmata 3.1.2 and 3.3.3. ⋈

In the following let the parameter of the Hopf-bifurcation λ be positive:

$$\lambda \;>\; 0.$$

We are only interested in supercritical bifurcations.

The restriction of the reduced vector field (cf. Lemma 3.2.1) to the statum $\Delta = 0$ is

$$\dot{\pi_1} = 2(\lambda + \alpha^r \pi_1 + \beta^r \pi_2)\pi_1 + 4\Big((\alpha-\beta)^r \pi_4 + (\alpha-\beta)^i \pi_5\Big) \tag{3.4.3}$$

$$\dot{\pi_2} = 2(\lambda + \beta^r \pi_1 + \alpha^r \pi_2)\pi_2 + 4\Big((\alpha-\beta)^r \pi_4 - (\alpha-\beta)^i \pi_5\Big) \tag{3.4.4}$$

$$\dot{\pi_3} = \frac{1}{2}\pi_2 \dot{\pi_2} \tag{3.4.5}$$

$$\begin{aligned}\dot{\pi_4} &= 2\Big(2\lambda + (\alpha+\beta)^r(\pi_1+\pi_2)\Big)\pi_4 + 2(\alpha-\beta)^i(-\pi_1+\pi_2)\pi_5 \\ &\quad +(\alpha-\beta)^r \pi_1\pi_2(\pi_1+\pi_2)\end{aligned} \tag{3.4.6}$$

$$\begin{aligned}\dot{\pi_5} &= 2\Big(2\lambda + (\alpha+\beta)^r(\pi_1+\pi_2)\Big)\pi_5 + 2(\alpha-\beta)^i(\pi_1-\pi_2)\pi_4 \\ &\quad +(\alpha-\beta)^i \pi_1\pi_2(-\pi_1+\pi_2).\end{aligned} \tag{3.4.7}$$

Here α^r, β^r resp. α^i, β^i denote the real resp. imaginary parts of α, β.

LEMMA 3.4.6 *Let $\alpha^r, \beta^r < 0$ and $\alpha^r \neq \beta^r$. Then the set of critical points of the Equations 3.4.3 to 3.4.7 on the stratum $\Delta = 0$ is given by a curve*

$$g(\pi_1) = \left(\pi_1, \pi_2 = -\left(\pi_1 + \frac{\lambda}{\alpha^r}\right), \frac{1}{4}\pi_2^2, \frac{1}{2}\pi_1\pi_2, 0\right),\ 0 \leq \pi_1 \leq -\frac{\lambda}{\alpha^r},$$

parametrised by π_1 and

$$h(\pi_1) = \left(\pi_1, \pi_1, \frac{1}{4}\pi_1^2, -\frac{1}{2}\pi_1^2, 0\right),\ \pi_1 = -\frac{\lambda}{2\beta^r}.$$

The curve $g(\pi_1)$, $0 \leq \pi_1 \leq -\frac{\lambda}{\alpha^r}$, connects a critical point with isotropy $(\mathrm{O}(2), 1)$,

$$g\left(-\frac{\lambda}{\alpha^r}\right) = \left(-\frac{\lambda}{\alpha^r}, 0, 0, 0, 0\right),$$

with a critical point with isotropy $(\mathrm{D}_2, \mathbb{Z}_2)$,

$$g(0) = \left(0, \pi_2 = -\frac{\lambda}{\alpha^r}, \frac{1}{4}\pi_2^2, 0, 0\right).$$

The critical point $h(\pi_1)$, $\pi_1 = -\frac{\lambda}{2\beta^r}$, lies in $\Pi(\mathrm{Fix}\,(\mathbb{T}, \mathbb{Z}_3))$, the image of points with isotropy $(\mathbb{T}, \mathbb{Z}_3)$ in the original system under the Hilbert-map.

Proof: By addition resp. subtraction of Equations 3.4.3 and 3.4.4 one gets the following equations

$$0 = \lambda(\pi_1+\pi_2) + \alpha^r(\pi_1^2+\pi_2^2) + 2\beta^r\pi_1\pi_2 + 4(\alpha-\beta)^r\pi_4, \tag{3.4.8}$$

$$0 = \lambda(\pi_1-\pi_2) + \alpha^r(\pi_1^2-\pi_2^2) + 4(\alpha-\beta)^i\pi_5. \tag{3.4.9}$$

Let $(\alpha-\beta)^i \neq 0$ then

$$\pi_4 = -\frac{\lambda(\pi_1+\pi_2) + \alpha^r(\pi_1^2+\pi_2^2) + 2\beta^r\pi_1\pi_2}{4(\alpha-\beta)^r},$$

$$\pi_5 = -\frac{\lambda(\pi_1-\pi_2) + \alpha^r(\pi_1^2-\pi_2^2)}{4(\alpha-\beta)^i}.$$

Inserting this in Equations 3.4.6 and 3.4.7 gives

$$\begin{aligned} 0 &= -\frac{(\pi_1+\pi_2)\Big(\lambda+\alpha^r(\pi_1+\pi_2)\Big)\Big(\lambda+\beta^r(\pi_1+\pi_2)\Big)}{(\alpha-\beta)^r}, \\ 0 &= -\frac{(\pi_1-\pi_2)\Big(\lambda+\alpha^r(\pi_1+\pi_2)\Big)}{2(\alpha-\beta)^r(\alpha-\beta)^i} \\ &\quad \Big(2\lambda(\alpha-\beta)^r+(\pi_1+\pi_2)(\alpha^{r2}-\beta^{r2}+(\alpha-\beta)^{i2}\Big). \end{aligned} \tag{3.4.10}$$

Looking for nontrivial critical points, one, therefore, has to study two cases.
Let $\pi_1+\pi_2=-\frac{\lambda}{\alpha^r}$. Since we assume $\lambda>0$, only the choice $\alpha^r<0$ gives solutions that lie in $\Pi(\mathbb{R}^6)$. By insertion one gets the curve

$$g(\pi_1)=\left(\pi_1,\pi_2=-\left(\pi_1+\frac{\lambda}{\alpha^r}\right),\frac{1}{4}\pi_2^2,\frac{1}{2}\pi_1\pi_2,0\right),\ 0\leq\pi_1\leq-\frac{\lambda}{\alpha^r},$$

of critical points. Lemma 3.4.1 gives the associated orbit types.
Now let $\pi_1+\pi_2=-\frac{\lambda}{\beta^r}$. Only the choice $\beta^r<0$ gives solutions that lie in $\Pi(\mathbb{R}^6)$ as above. By insertion in Equation 3.4.10 one gets the condition

$$0 = \frac{\Big((\alpha-\beta)^{r2}+(\alpha-\beta)^{i2}\Big)\lambda^2(\lambda+2\beta^r\pi_2)}{2\beta^{r3}(\alpha-\beta)^i}.$$

In order to get critical points, one has to choose

$$\pi_1=\pi_2=-\frac{\lambda}{2\beta^r}.$$

By insertion one obtains the critical point

$$h(\pi_1)=\left(\pi_1,\pi_1,\frac{1}{4}\pi_1^2,-\frac{1}{2}\pi_1^2,0\right),\ \pi_1=-\frac{\lambda}{2\beta^r},$$

lying in $\Pi(\mathrm{Fix}(\mathbb{T},\mathbb{Z}_3))$ (cf. Lemma 3.4.5). It shall be shown that there are no other critical points with radius

$$\pi_1+\pi_2 = -\frac{\lambda}{\beta^r}.$$

Therefore the group orbit of periodic orbits with isotropy $(\mathbb{T},\mathbb{Z}_3)$ in the original system can only intersect the stratified space in the curve given in Lemma 3.4.5.
Now let $(\alpha-\beta)^i=0$. Equations 3.4.8 and 3.4.9 yield

$$\begin{aligned} 0 &= \lambda(\pi_1+\pi_2)+\alpha^r(\pi_1^2+\pi_2^2)+2\beta^r\pi_1\pi_2+4(\alpha-\beta)^r\pi_4, \\ 0 &= (\pi_1-\pi_2)\Big(\lambda+\alpha^r(\pi_1+\pi_2)\Big). \end{aligned}$$

Consequently we have to study two cases.
Let $\pi_1=\pi_2$. Then

$$\pi_4 = -\frac{\Big(\lambda+(\alpha+\beta)^r\pi_1\Big)\pi_1}{2(\alpha-\beta)^r}.$$

By insertion in Equation 3.4.6 one gets

$$0 = -\frac{\pi_1(\lambda + 2\beta^r\pi_1)(\lambda + 2\alpha^r\pi_1)}{(\alpha-\beta)^r}.$$

The choice $\pi_1 = \pi_2 = -\frac{\lambda}{2\alpha^r}$ and the relation

$$\pi_4^2 + \pi_5^2 = \pi_1^2\pi_3 = \frac{1}{4}\pi_1^4$$

give the critical point

$$\left(\pi_1, \pi_1, \frac{1}{4}\pi_1^2, \frac{1}{2}\pi_1^2, 0\right), \ \pi_1 = -\frac{\lambda}{2\alpha^r},$$

that lies on the curve $g(\pi_1)$.
The case $\pi_1 = \pi_2 = -\frac{\lambda}{2\beta^r}$ again yields the solution $h(\pi_1)$, $\pi_1 = -\frac{\lambda}{2\beta^r}$.
Finally we have to study the case $\pi_1 + \pi_2 = -\frac{\lambda}{\alpha^r}$. We get

$$\begin{aligned}\pi_4 &= -\frac{\pi_1}{2\alpha^r}(\lambda + \alpha^r\pi_1)\\ &= \frac{1}{2}\pi_1\pi_2.\end{aligned}$$

The relation

$$\pi_4^2 + \pi_5^2 = \frac{1}{4}\pi_1^2\pi_2^2$$

yields $\pi_5 = 0$. So again we get the curve $g(\pi_1)$. ⋈

LEMMA 3.4.7

$$\Pi(\mathrm{Fix}\,(\mathbb{T}, \mathbb{Z}_3)) \cap g(\pi_1) = \emptyset, \ 0 \le \pi_1 \le -\frac{\lambda}{\alpha^r}.$$

The critical point $h(\pi_1)$, $\pi_1 = -\frac{\lambda}{2\beta^r}$, (cf. Lemma 3.4.6) that lies in $\Pi(\mathrm{Fix}(\mathbb{T}, \mathbb{Z}_3))$ is isolated in the Hilbert-set $\Pi(\mathbb{R}^6)$.

Proof: For points lying on the curve $g(\pi_1)$ we have $\pi_1 + \pi_2 = -\frac{\lambda}{\alpha^r}$. Points in $\Pi(\mathrm{Fix}\,(\mathbb{T}, \mathbb{Z}_3))$ satisfy the condition $\pi_1 = \pi_2$ (cf. Lemma 3.4.5). For a potential intersection this means $\pi_1 = \pi_2 = -\frac{\lambda}{2\alpha^r}$. We have

$$g\left(-\frac{\lambda}{2\alpha^r}\right) = \left(-\frac{\lambda}{2\alpha^r}, -\frac{\lambda}{2\alpha^r}, \frac{1}{16}\frac{\lambda^2}{\alpha^{r2}}, +\frac{1}{8}\frac{\lambda^2}{\alpha^{r2}}, 0\right)$$

whereas

$$\Pi(\mathrm{Fix}\,(\mathbb{T}, \mathbb{Z}_3)) \cap \left(\pi_1 = -\frac{\lambda}{2\alpha^r}\right) = \left(-\frac{\lambda}{2\alpha^r}, -\frac{\lambda}{2\alpha^r}, \frac{1}{16}\frac{\lambda^2}{\alpha^{r2}}, -\frac{1}{8}\frac{\lambda^2}{\alpha^{r2}}, 0\right).$$

On the stratum $\Delta = 0$ the critical point $h(\pi_1)$, $\pi_1 = -\frac{\lambda}{2\beta^r}$, (cf. Lemma 3.4.6) that lies on $\Pi(\mathrm{Fix}(\mathbb{T}, \mathbb{Z}_3))$, therefore, is isolated. We shall show in Lemma 3.4.8 that there are no further critical points in the Hilbert-set in the region $\Delta \neq 0$ near $h(\pi_1)$, $\pi_1 = -\frac{\lambda}{2\beta^r}$. ⋈

Now we are looking for critical points of the reduced vector field (cf. Lemma 3.2.1) in $\Pi(\mathbb{R}^6)$ that do not lie on the stratum $\Delta = 0$. Such a critical point has to meet the condition (cf. Lemma 3.4.4)

$$\dot{\Delta} = 4\,\Delta\ (\lambda + \beta^r \pi_1 + \gamma^r \pi_2) = 0.$$

Since we assumed $\Delta \neq 0$, this means

$$\lambda + \beta^r \pi_1 + \gamma^r \pi_2 \quad = \quad 0. \tag{3.4.11}$$

So we get the following equations:

$$\begin{aligned}
0 &= 2(\lambda + \alpha^r \pi_1 + \beta^r \pi_2)\pi_1 + 4\Big((\alpha-\beta)^r \pi_4 + (\alpha-\beta)^i \pi_5\Big) && (3.4.12)\\
0 &= 8(\alpha-\gamma)^r \pi_3 + 4\Big((\alpha-\beta)^r \pi_4 - (\alpha-\beta)^i \pi_5\Big) && (3.4.13)\\
0 &= \frac{1}{2}\pi_2 \dot{\pi}_2 && (3.4.14)\\
0 &= 2\Big(2\lambda + (\alpha+\beta)^r(\pi_1+\pi_2)\Big)\pi_4 + 2(\alpha-\beta)^i(-\pi_1+\pi_2)\pi_5 \\
&\quad +(\alpha-\beta)^r \pi_1(\pi_1\pi_2 + 4\pi_3) && (3.4.15)\\
0 &= 2\Big(2\lambda + (\alpha+\beta)^r(\pi_1+\pi_2)\Big)\pi_5 + 2(\alpha-\beta)^i(\pi_1-\pi_2)\pi_4 \\
&\quad +(\alpha-\beta)^i \pi_1(-\pi_1\pi_2 + 4\pi_3) && (3.4.16)\\
\pi_2 &= -\frac{\lambda + \beta^r \pi_1}{\gamma^r}. && (3.4.17)
\end{aligned}$$

Here $\alpha^r, \beta^r, \gamma^r$ resp. α^i, β^i again denote the real resp. imaginary parts of α, β, γ. In the following we shall assume

$$\beta^r < \alpha^r < \gamma^r < 0.$$

In Lemma 3.5.1 we shall show that only for this choice of the coefficients the solutions with isotropy $(\mathrm{O}(2), 1)$ resp. $\widetilde{\mathrm{SO}(2)}$ can be stable simultaneously. Investigations using the topological Conlex-index suggested to study this case. In the following lemma the solution with isotropy $\widetilde{\mathrm{SO}(2)}$ is being described.

Lemma 3.4.8 *Let $\beta^r < \alpha^r < \gamma^r < 0$. Then*

$$\left(0, -\frac{\lambda}{\gamma^r}, 0, 0, 0\right)$$

is the only critical point of the reduced vector field in $\Pi(\mathbb{R}^6)$ with $\Delta \neq 0$. This solution has isotropy $\widetilde{\mathrm{SO}(2)}$.

Proof: First let $(\alpha-\beta)^i \neq 0$. By addition resp. subtraction of Equations 3.4.12 and 3.4.13 we get

$$\begin{aligned}
0 &= (\lambda + \alpha^r \pi_1 + \beta^r \pi_2)\pi_1 + 4(\alpha-\beta)^r \pi_4 + 4(\alpha-\gamma)^r \pi_3,\\
0 &= (\lambda + \alpha^r \pi_1 + \beta^r \pi_2)\pi_1 + 4(\alpha-\beta)^i \pi_5 - 4(\alpha-\gamma)^r \pi_3.
\end{aligned}$$

Therefore we have

$$\begin{aligned}\pi_4 &= -\frac{(\lambda+\alpha^r\pi_1+\beta^r\pi_2)\pi_1+4(\alpha-\gamma)^r\pi_3}{4(\alpha-\beta)^r},\\ \pi_5 &= -\frac{(\lambda+\alpha^r\pi_1+\beta^r\pi_2)\pi_1-4(\alpha-\gamma)^r\pi_3}{4(\alpha-\beta)^i}.\end{aligned}$$

Insertion in Equation 3.4.15 yields

$$\frac{(\beta-\gamma)^r\Big(\lambda(\alpha-\gamma)^r+\alpha^r\pi_1(\beta-\gamma)^r\Big)(-\lambda\pi_1-\beta^r\pi_1^2+4\gamma^r\pi_3)}{(\alpha-\beta)^r\gamma^{r2}} = 0.$$

Let

$$\pi_1=-\frac{\lambda(\alpha-\gamma)^r}{\alpha^r(\beta-\gamma)^r}>0.$$

Using Equation 3.4.17 we get

$$\pi_2=\frac{\lambda(\alpha-\beta)^r}{\alpha^r(\beta-\gamma)^r}>0.$$

Together with Equation 3.4.16 this yields

$$0 = \frac{(\alpha-\gamma)^r\Big((\alpha-\beta)^{r2}+(\alpha-\beta)^{i2}\Big)\lambda\Big(-\lambda^2(\alpha-\beta)^{r2}+4\alpha^{r2}\pi_3(\beta-\gamma)^{r2}\Big)}{2\alpha^{r3}(\alpha-\beta)^r(\alpha-\beta)^i(\beta-\gamma)^{r2}}.$$

So we have

$$\pi_3=\frac{\lambda^2(\alpha-\beta)^{r2}}{4\alpha^{r2}(\beta-\gamma)^{r2}}=\frac{1}{4}\pi_2^2.$$

This solution lies on the stratum $\Delta=0$.
Now let

$$\pi_3=\frac{\pi_1(\lambda+\beta^r\pi_1)}{4\gamma^r}=-\frac{1}{4}\pi_1\pi_2.$$

Insertion in Equation 3.4.16 yields

$$\begin{aligned}0 &= -2(\alpha-\beta)^{i2}\pi_1^2\pi_2-\frac{(\alpha-\beta)^{i2}\pi_1\Big(\lambda+(\beta+\gamma)^r\pi_1\Big)^2}{2\gamma^{r2}}\\ &\quad -\frac{\pi_1\Big(\lambda(\alpha+\beta-2\gamma)^r+\pi_1(\alpha+\beta)^r(\beta-\gamma)^r\Big)^2}{2\gamma^{r2}}.\end{aligned}$$

Since all elements of the sum are nonpositive in $\Pi(\mathbb{R}^6)$ the sum can only be zero if all elements are zero. This is only possible if $\pi_1=0$. This yields

$$\left(0,-\frac{\lambda}{\gamma^r},0,0,0\right),$$

the solution with isotropy $\widetilde{\mathrm{SO}(2)}$.

Second let $(\alpha-\beta)^i=0$. Again by addition resp. subtraction of the Equations 3.4.12 and 3.4.13 we get

$$\begin{aligned}\pi_3 &= \frac{(\lambda+\alpha^r\pi_1+\beta^r\pi_2)\pi_1}{4(\alpha-\gamma)^r},\\ \pi_4 &= -\frac{2(\alpha-\gamma)^r\pi_3}{(\alpha-\beta)^r}.\end{aligned} \tag{3.4.18}$$

Furthermore we have

$$\pi_2 = -\frac{\lambda+\beta^r\pi_1}{\gamma^r}.$$

Insertion in Equation 3.4.16 yields

$$0 = \Big(2\lambda+(\alpha+\beta)^r(\pi_1+\pi_2)\Big)\pi_5.$$

In order to solve this equation, we have to look at several cases.
Let $\pi_5=0$. Then

$$\pi_4^2+\pi_5^2 = \pi_1^2\pi_3$$

and Equation 3.4.18 yields

$$\frac{4(\alpha-\gamma)^{r2}}{(\alpha-\beta)^{r2}}\pi_3^2 = \pi_1^2\pi_3.$$

For $\pi_3\neq 0$ we get

$$\pi_1=-\frac{\lambda(\alpha-\gamma)^r}{\alpha^r(\beta-\gamma)^r}>0,\ \pi_2=\frac{\lambda(\alpha-\beta)^r}{\alpha^r(\beta-\gamma)^r}>0,\ \pi_3=\frac{1}{4}\pi_2^2.$$

Therefore the solution lies on the stratum $\Delta=0$.
The choice $\pi_3=0$ and Equation 3.4.16 yield the solution

$$\left(0,-\frac{\lambda}{\gamma^r},0,0,0\right)$$

with isotropy $\widetilde{\mathrm{SO}(2)}$. For $\pi_5\neq 0$ and

$$0 = 2\lambda+(\alpha+\beta)^r(\pi_1+\pi_2)$$

Equation 3.4.15 gives

$$0 = \pi_1(\pi_1\pi_2+4\pi_3).$$

Choosing $\pi_1=0$ again yields the solution with isotropy $\widetilde{\mathrm{SO}(2)}$.
For $\pi_1\neq 0$ one gets the solution

$$\begin{aligned}\pi_1 &= -\frac{\lambda(\alpha+\beta-2\gamma)^r}{(\alpha+\beta)^r(\beta-\gamma)^r}>0,\\ \pi_2 &= \frac{\lambda(\alpha-\beta)^r}{(\alpha+\beta)^r(\beta-\gamma)^r}>0,\\ \pi_3 &= \frac{\lambda^2(\alpha-\beta)^r(\alpha+\beta-2\gamma)^r}{4(\alpha+\beta)^{r2}(\beta-\gamma)^{r2}}<0\end{aligned}$$

that does not lie in $\Pi(\mathbb{R}^6)$. $\bowtie$

Figure 4 sketches the position of the critical points of the reduced vector field (cf. Lemma 3.2.1) in the Hilbert-set $\Pi(\mathbb{R}^6)$ known by Lemmata 3.4.6 and 3.4.8 under the assumption

$$\beta^r < \alpha^r < \gamma^r < 0.$$

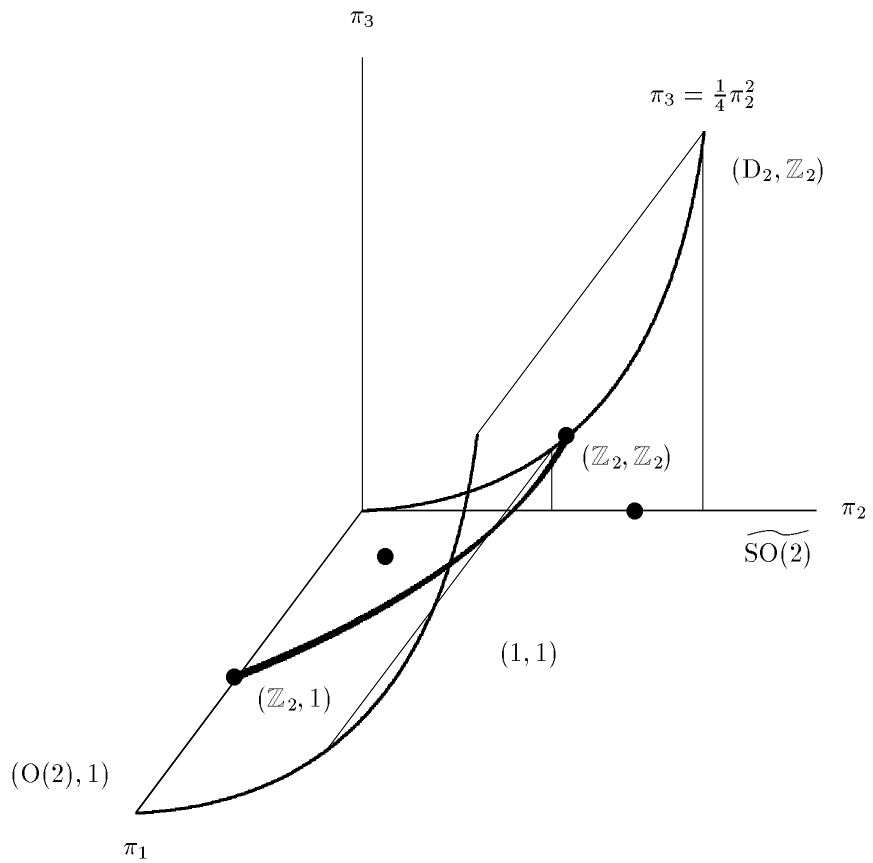


Figure 4: Critical points of the reduced vector field in the Hilbert-set

We now study the curve

$$g(\pi_1) = \left(\pi_1, \pi_2 = -\left(\pi_1 + \frac{\lambda}{\alpha^r}\right), \frac{1}{4}\pi_2^2, \frac{1}{2}\pi_1\pi_2, 0\right), \quad 0 < \pi_1 < -\frac{\lambda}{\alpha^r},$$

of critical points of the Equations (3.4.3) to (3.4.7) (cf. Lemma 3.4.6).

LEMMA 3.4.9 *The preimage of a point $g(\pi_1), \pi_1 \in \left(0, -\frac{\lambda}{\alpha^r}\right)$, in $\mathbb{R}^6$ is a two-torus. It is fibered with periodic solutions.*

Proof: The curve g of critical points lies on the statum

$$\Delta = \frac{1}{4}\pi_2^2 - \pi_3 = 0.$$

Introducing polar coordinates in the form

$$z_j = r_j \, e^{\mathrm{i}\phi_j}, \quad j \in \{-2, 0, 2\},$$

yields

$$r_{-2} \;=\; r_2.$$

The choice of

$$\pi_1 = r_0^2 \in \left(0, -\frac{\lambda}{\alpha^r}\right)$$

and the condition

$$\pi_1 + \pi_2 \;=\; -\frac{\lambda}{\alpha^r}$$

determine the radii. Let

$$\theta \;=\; 2\phi_0 - \phi_{-2} - \phi_2.$$

Then the conditions for π_4 resp. π_5 yield, in polar coordinates, the phase relations

$$\cos\theta = 1, \quad \sin\theta = 0$$

and, thus,

$$\theta \;=\; 0 \bmod 2\pi.$$

So one angle is determined, two are still available, the preimage is a 2-torus. Points on the surface $\Delta = 0$ have the (conjugated) isotropy $(\mathbb{Z}_2, 1)$. Therefore it is possible just to look at points of the form (z_2, z_0, z_2) in order to determine the resulting flow on the preimage of a point on the curve of fixed points. Thus we have the additional condition

$$\phi_{-2} \;=\; \phi_2.$$

Using $\theta = 0 \bmod 2\pi$, one sees that

$$\phi_0 \;=\; \phi_2 \bmod \pi.$$

Inserting this into the differential equation yields

$$\dot{\phi}_0 = \omega_0 = \omega + \alpha^i \left(r_0^2 + 2r_2^2\right).$$

Thus the $2-$torus is fibered with periodic solutions of period near $\frac{2\pi}{\omega}$. ⋈

3.5 Stability of the critical points of the reduced vector field

In Lemmata 3.4.6 and 3.4.8 we have shown that in the case of supercritical bifurcation $(\lambda > 0)$ the coefficients $\alpha^r, \beta^r, \gamma^r$ have to be negative in order that the corresponding solutions lie in the image of the Hilbert-map $\Pi(\mathbb{R}^6)$. The following lemma gives a condition on the choice of the coefficients relative to each other.

LEMMA 3.5.1 *Only by choosing the coefficients*

$$\beta^r < \alpha^r < \gamma^r < 0, \quad \alpha^r \in \left(\frac{1}{4}(\beta+3\gamma)^r, \gamma^r\right),$$

the critical points with isotropy $(\mathrm{O}(2),1)$ *resp.* $\widetilde{\mathrm{SO}(2)}$ *of the reduced vector field (cf. Lemma 3.2.1) can be stable simultaneously. The stability of the solution with isotropy* $(\mathrm{O}(2),1)$ *is determined by higher order terms because of the existence of a curve of critical points (cf. Lemma 3.4.6).*

Proof: The calculations of [IoRo] yield (using our parameters) up to third order the following conditions for the stability of the periodic solutions with isotropy $(\mathrm{O}(2),1)$ resp. $\widetilde{\mathrm{SO}(2)}$ in the original, ten dimensional system:

isotropy	nontrivial Floquet-exponents
$(\mathrm{O}(2),1)$	$-2\lambda < 0, -\frac{2\lambda}{\alpha^r}(\alpha-\beta)^r < 0, -\frac{2\lambda}{\alpha^r}(-4\alpha+\beta+3\gamma)^r < 0$
$\widetilde{\mathrm{SO}(2)}$	$-2\lambda < 0, -\frac{2\lambda}{\gamma^r}(\alpha-\gamma), cc, \frac{\lambda}{\gamma^r}(\gamma-\beta), cc, \frac{3\lambda}{2\gamma^r}(\gamma-\beta), cc.$

Here cc denotes the complex conjugate of the preceding number.
So we get the conditions

$$\beta^r < \alpha^r < \gamma^r < 0$$

and

$$\beta^r + 3\gamma^r < 4\alpha^r.$$

The ansatz

$$\alpha^r \;=\; t\,\beta^r + (1-t)\,\gamma^r, \quad t \in (0,1),$$

yields

$$(\beta-\gamma)^r(1-4t) \;<\; 0$$

and, therefore, we have

$$t \;\in\; \left(0, \frac{1}{4}\right).$$

This means

$$\alpha^r \;\in\; \left(\frac{1}{4}(\beta+3\gamma)^r, \gamma^r\right).$$

Especially

$$\frac{(\alpha-\gamma)^r}{(\beta-\gamma)^r} \;\in\; \left(0, \frac{1}{4}\right). \qquad \bowtie$$

Now we want to determine the linearization of the reduced vector field (cf. Lemma 3.2.1) along the curve $g(\pi_1)$ of critical points (cf. Lemma 3.4.6). For the general linearization L one gets

$$L = \begin{pmatrix} 2\lambda + 4\pi_1\alpha^r + 2\pi_2\beta^r & 2\pi_1\beta^r & 0 & 4(\alpha-\beta)^r & 4(\alpha-\beta)^i \\ 2\pi_2\beta^r & 2\lambda + 2\pi_1\beta^r + 4\pi_2\gamma^r & 8(\alpha-\gamma)^r & 4(\alpha-\beta)^r & -4(\alpha-\beta)^i \\ 4\pi_3\beta^r & 4\pi_3\alpha^r + 2\big(\pi_4(\alpha-\beta)^r - \pi_5(\alpha-\beta)^i\big) & 4(\lambda + \beta^r\pi_1 + \alpha^r\pi_2) & 2\pi_2(\alpha-\beta)^r & -2\pi_2(\alpha-\beta)^i \\ 2\pi_4(\alpha+\beta)^r - 2\pi_5(\alpha-\beta)^i + (2\pi_1\pi_2 + 4\pi_3)(\alpha-\beta)^r & 2\pi_4(\alpha+\beta)^r + 2\pi_5(\alpha-\beta)^i + \pi_1^2(\alpha-\beta)^r & 4\pi_1(\alpha-\beta)^r & 4\lambda + 2(\pi_1+\pi_2)(\alpha+\beta)^r & 2(-\pi_1+\pi_2)(\alpha-\beta)^i \\ 2\pi_5(\alpha+\beta)^r + 2\pi_4(\alpha-\beta)^i + (-2\pi_1\pi_2 + 4\pi_3)(\alpha-\beta)^i & 2\pi_5(\alpha+\beta)^r - 2\pi_4(\alpha-\beta)^i - \pi_1^2(\alpha-\beta)^i & 4\pi_1(\alpha-\beta)^i & 2(\pi_1-\pi_2)(\alpha-\beta)^i & 4\lambda + 2(\pi_1+\pi_2)(\alpha+\beta)^r \end{pmatrix}.$$

We are interested in the eigenvalues of L along the curve $g(\pi_1)$ with reference to $\Pi(\mathbb{R}^6) \subset \mathbb{R}^5$. Thus we have to determine the tangent space at points of the curve in $\Pi(\mathbb{R}^6)$. It is given by the relation

$$\pi_4^2 + \pi_5^2 \;=\; \pi_1^2\pi_3.$$

The curve itself lies on the stratum

$$\Delta = \frac{1}{4}\pi_2^2 - \pi_3 = 0.$$

So we get the following lemma.

LEMMA 3.5.2 *The tangent space at the stratum $\Delta = 0$ along the curve*

$$g(\pi_1),\; 0 < \pi_1 < -\frac{\lambda}{\alpha^r},$$

is spanned by the vectors

$$\begin{aligned} t_1 &= \left(1, -1, -\frac{1}{2}\pi_2, \frac{1}{2}(\pi_2 - \pi_1), 0\right), \\ t_2 &= \left(\pi_1, \pi_2, \frac{1}{2}\pi_2^2, \pi_1\pi_2, 0\right), \\ t_3 &= \big(2\alpha^r(\alpha-\beta)^i, -2\alpha^r(\alpha-\beta)^i, -\pi_2\alpha^r(\alpha-\beta)^i, \alpha^r(\pi_2-\pi_1)(\alpha-\beta)^i, \\ &\quad -\alpha^r(\pi_1+\pi_2)(\alpha-\beta)^r\big). \end{aligned}$$

The vectors t_1, t_2, t_3 are eigenvectors of L to the eigenvalues

$$\begin{aligned} ew_1 &= 0, \\ ew_2 &= -2\lambda \;=\; 2\alpha^r(\pi_1+\pi_2), \\ ew_3 &= \frac{2(\alpha-\beta)^r\lambda}{\alpha^r} \;=\; -2(\alpha-\beta)^r(\pi_1+\pi_2). \end{aligned}$$

The curve $g(\pi_1)$, $0<\pi_1<-\frac{\lambda}{\alpha^r}$, *is stable on the stratum* $\Delta=0$.

Proof: The relations $\pi_4^2+\pi_5^2=\pi_1^2\pi_3$ and $\Delta=0$ yield the following vectors normal to the tangent space at the surface $\Delta=0$ in $\Pi(\mathbb{R}^6)\subset\mathbb{R}^5$:

$$\begin{aligned} n_1 &= \left(-2\pi_1\pi_3, 0, -\pi_1^2, 2\pi_4, 2\pi_5\right), \\ n_2 &= \left(0, \frac{1}{2}\pi_2, -1, 0, 0\right). \end{aligned}$$

The orthogonal complement to $Span(n_1,n_2)$ is spanned by the vectors t_1,t_2,t_3. A simple calculation shows that these vectors are eigenvectors to the given eigenvalues. The eigenvector t_1 points along the curve of critical points. Therefore the associated eigenvalue is zero. By definition of the curve $g(\pi_1)$ we have

$$\pi_1+\pi_2=-\frac{\lambda}{\alpha^r}.$$

Therefore the curve $g(\pi_1)$, $0<\pi_1<-\frac{\lambda}{\alpha^r}$, is stable on the stratum $\Delta=0$. $\bowtie$

Now we want to determine the linearization of the reduced vector field along the curve $g(\pi_1)$ of fixed points in the direction of the principal stratum. We shall show that there exists a point $\tilde{\pi}_1$ on the curve $g(\pi_1)$ in which the stability of the curve changes from stable to unstable in the direction of the principal stratum. In this point the linearization L of the vector field of the reduced equation has a nontrivial two dimensional Jordan-block with respect to the eigenvalue zero.
Let

$$t = (0,1,0,0,0).$$

Then $n_1t=0$ and $n_2t\neq 0$ for $\pi_2\neq 0$. Thus the vectors t_1,t_2,t_3,t span the tangent space at the Hilbert-set $\Pi(\mathbb{R}^6)$ along the curve $g(\pi_1)$, $0<\pi_1<-\frac{\lambda}{\alpha^r}$. One gets

$$Lt = a\,t_1+b\,t_2+c\,t_3+d\,t$$

with

$$\begin{aligned} a &= -2\pi_1\frac{(\alpha-\beta)^{r2}+(\alpha-\beta)^{i2}}{(\alpha-\beta)^r}, \\ b &= 2\alpha^r, \\ c &= \pi_1\frac{(\alpha-\beta)^i}{\alpha^r(\alpha-\beta)^r}, \\ d &= 4\frac{\lambda(\alpha-\gamma)^r+\pi_1\alpha^r(\beta-\gamma)^r}{\alpha^r}. \end{aligned}$$

Restricted to the tangent space at the curve $g(\pi_1)$, $0<\pi_1<-\frac{\lambda}{\alpha^r}$, according to our choice of the vectors t_1,t_2,t_3,t, L has the form $\tilde{L}$:

$$\tilde{L} = \begin{pmatrix} 0 & 0 & 0 & a \\ 0 & -2\lambda & 0 & b \\ 0 & 0 & \frac{2(\alpha-\beta)^r\lambda}{\alpha^r} & c \\ 0 & 0 & 0 & d \end{pmatrix}.$$

Especially the fourth eigenvalue is

$$\begin{aligned} ew_4 &= d \\ &= 4\frac{\lambda(\alpha-\gamma)^r + \pi_1\alpha^r(\beta-\gamma)^r}{\alpha^r} = -4\Big(\pi_1(\alpha-\beta)^r + \pi_2(\alpha-\gamma)^r\Big). \end{aligned}$$

For

$$\tilde{\pi}_1 = -\frac{\lambda(\alpha-\gamma)^r}{\alpha^r(\beta-\gamma)^r}$$

we have $ew_4(\tilde{\pi}_1) = 0$. Choosing the coefficients according to Lemma 3.5.1 yields

$$0 < \tilde{\pi}_1 < -\frac{\lambda}{4\alpha^r}.$$

The point $g(\tilde{\pi}_1)$ is exactly the intersection point of the curve $g(\pi_1)$, $0 \leq \pi_1 \leq -\frac{\lambda}{\alpha^r}$, with the surface $\lambda + \beta^r\pi_1 + \gamma^r\pi_2 = 0$ (cf. Lemma 3.4.4). Only on this surface we can have critical points of the reduced vector field (cf. Lemma 3.2.1) outside the stratum $\Delta = 0$ (cf. Lemma 3.4.6).

For

$$h(\tilde{\pi}_1) = \begin{pmatrix} -\frac{(\alpha-\gamma)^r\lambda\left((\alpha-\beta)^{r2}-(\alpha-\beta)^{i2}\right)}{(\alpha-\beta)^{r2}(\beta-\gamma)^r} \\ \frac{(\alpha-\gamma)^r\lambda\left((\alpha-\beta)^{r2}-(\alpha-\beta)^{i2}\right)}{(\alpha-\beta)^{r2}(\beta-\gamma)^r} \\ \frac{\lambda^2\left((\alpha-\beta)^{r3}-(\alpha-\gamma)^r(\alpha-\beta)^{i2}\right)}{2\alpha^r(\alpha-\beta)^r(\beta-\gamma)^{r2}} \\ -\frac{(\alpha-\gamma)^r\lambda^2\left(2(\alpha-\beta)^{r3}+(-2\alpha+\beta+\gamma)^r(\alpha-\beta)^{i2}\right)}{2\alpha^r(\alpha-\beta)^{r2}(\beta-\gamma)^{r2}} \\ \frac{(\alpha-\gamma)^r(\alpha-\beta)^i\lambda^2}{2\alpha^r(\alpha-\beta)^r(\beta-\gamma)^r} \end{pmatrix}$$

we have

$$h(\tilde{\pi}_1) = \alpha^r t_2(\tilde{\pi}_1) - \frac{(\alpha-\beta)^i\tilde{\pi}_1}{2(\alpha-\beta)^{r2}}t_3(\tilde{\pi}_1) + \lambda t.$$

Consequently, $h(\tilde{\pi}_1) \in \operatorname{Span}\Big(t_1(\tilde{\pi}_1), t_2(\tilde{\pi}_1), t_3(\tilde{\pi}_1), t\Big)$, and one sees that $L\,h(\tilde{\pi}_1) = j\,t_1(\tilde{\pi}_1)$ with

$$j = \frac{2\lambda^2(\alpha-\gamma)^r\Big((\alpha-\beta)^{r2} + (\alpha-\beta)^{i2}\Big)}{\alpha^r(\alpha-\beta)^r(\beta-\gamma)^r} < 0.$$

So we have shown the following lemma.

LEMMA 3.5.3 *In the point $g(\tilde{\pi}_1)$, $\tilde{\pi}_1 = -\frac{\lambda(\alpha-\gamma)^r}{\alpha^r(\beta-\gamma)^r}$, the linearization L of the vector field of the reduced equation (cf. Lemma 3.2.1) has a nontrivial, two dimensional Jordan-block with respect to the eigenvalue zero.*

Up to now we have studied the reduced vector field resulting from the normal form up to third order (cf. [IoRo]). It has been shown in this section that this vector field is degenerate. In the next section we shall use fifth order terms to investigate this degeneracy.

3.6 Fifth order terms

Restricted to $\mathrm{Fix}(\mathbb{Z}_2, 1)$ the normal form (cf. [IoRo]) yields the following fifth order terms (FOT). Proceed as in Chapter 3.2 to get these terms.

$$\begin{aligned}\mathrm{FOT} &= (\delta_1\pi_1^2 + \delta_2\pi_1\pi_2 + \delta_3\pi_2^2 + \delta_4\pi_3 + \delta_5\pi_4)\, e_1 \\ &\quad + (\delta_6\pi_1^2 + \delta_7\pi_1\pi_2 + \delta_8\pi_2^2 + \delta_9\pi_3 + \delta_{10}\pi_4 + \delta_{11} i\pi_5)\, e_2 \\ &\quad + (\delta_{12}\pi_1 + \delta_{13}\pi_2)\, e_3 + (\delta_{14}\pi_1 + \delta_{15}\pi_2)\, e_4 - \frac{1}{2}\delta_{11}\pi_2 i e_5.\end{aligned}$$

The coefficients $\delta_1, \ldots, \delta_{15} \in \mathbb{C}$ result from a transformation of the coefficients $d_1, \ldots, d_9 \in \mathbb{C}$ of the normalform (cf. [IoRo])

$$\begin{aligned} d_3 &\to \sqrt{6}\, d_3 \\ d_4 &\to -d_4 \\ d_5 &\to -\sqrt{6}\, d_5 \\ d_6 &\to \sqrt{\tfrac{3}{2}}\, d_6 \\ d_7 &\to \sqrt{\tfrac{3}{2}}\, d_7 \\ d_9 &\to \tfrac{3}{8}\, d_9 \end{aligned}$$

as follows

$$\begin{aligned}
\delta_1 &= d_1 + \frac{1}{4}d_2 - 3d_3 + \frac{1}{2}d_4 - \frac{3}{2}d_5 + d_6 + d_7 - d_8 \\
\delta_2 &= 2d_1 - 7d_3 + \frac{1}{2}d_4 - 2d_5 - 2d_7 + 3d_8 \\
\delta_3 &= d_1 - 4d_3 - 2d_8 + d_9 \\
\delta_4 &= d_2 + 2d_5 + 12d_6 + 4d_7 - 4d_9 \\
\delta_5 &= d_2 + 2d_5 - 4d_6 - 4d_7 \\
\delta_6 &= d_1 + \frac{1}{4}d_2 - 4d_3 + \frac{1}{2}d_5 - 2d_7 + d_8 \\
\delta_7 &= 2d_1 - 4d_3 + 4d_7 - 2d_8 \\
\delta_8 &= d_1 \\
\delta_9 &= d_2 - 6d_5 \\
\delta_{10} &= d_2 - 2d_5 + 4d_7 - 2d_8 \\
\delta_{11} &= -4d_5 + 4d_7 + 2d_8 \\
\delta_{12} &= -6d_3 + d_4 - 4d_5 + 12d_6 - 2d_8 \\
\delta_{13} &= -6d_3 + d_4 \\
\delta_{14} &= 2d_3 + d_4 - 4d_5 - 4d_6 + 2d_8 \\
\delta_{15} &= 2d_3 + d_4 - 2d_5 + 2d_7 - 3d_8.
\end{aligned}$$

In the following we want to study the vector field perturbed in fifth order of the form

$$\dot{\pi} = f(\pi) + \varepsilon\, \mathrm{RFOT}(\pi), \;\; \varepsilon \ll 1.$$

By reduction of the fifth order terms (FOT) to the orbit space one gets the perturbation RFOT (reduced fifth order terms) with components RFOT1, ..., RFOT5.

$$\begin{aligned}
\mathrm{RFOT1} &= 2\pi_1(\delta_1^r\pi_1^2+\delta_2^r\pi_1\pi_2+\delta_3^r\pi_2^2+\delta_4^r\pi_3)+\pi_4\Big(2(\delta_5+\delta_{14})^r\pi_1\\
&\quad +(\delta_{11}+2\delta_{15})^r\pi_2\Big)+\pi_5\Big(2\delta_{14}^i\pi_1+(\delta_{11}+2\delta_{15})^i\pi_2\Big)\\
\mathrm{RFOT2} &= 4\delta_{12}^r\pi_1\pi_3+2\pi_2\Big(\delta_6^r\pi_1^2+\delta_7^r\pi_1\pi_2+\delta_8^r\pi_2^2+(\delta_9+2\delta_{13})^r\pi_3\Big)\\
&\quad +\pi_4\Big(2\delta_{14}^r\pi_1+(2\delta_{10}-\delta_{11}+2\delta_{15})^r\pi_2\Big)-\pi_5\Big(2\delta_{14}^i\pi_1\\
&\quad +(\delta_{11}+2\delta_{15})^i\pi_2\Big)\\
\mathrm{RFOT3} &= 2\pi_3\Big(2\delta_6^r\pi_1^2+(2\delta_7+\delta_{12})^r\pi_1\pi_2+(2\delta_8+\delta_{13})^r\pi_2^2+2\delta_9^r\pi_3\Big)\\
&\quad +\pi_4\left(\pi_2\left(\delta_{14}^r\pi_1-\left(\frac{1}{2}\delta_{11}-\delta_{15}\right)^r\pi_2\right)+4\delta_{10}^r\pi_3\right)\\
&\quad -\pi_5\left(\pi_2\left(\delta_{14}^i\pi_1-\left(\frac{1}{2}\delta_{11}-\delta_{15}\right)^i\pi_2\right)+4\delta_{11}^i\pi_3\right)\\
\mathrm{RFOT4} &= 2\pi_4\left((\delta_1+\delta_6)^r\pi_1^2+\left(\delta_2+\delta_7+\frac{1}{2}\delta_{12}\right)^r\pi_1\pi_2\right.\\
&\quad \left.+\left(\delta_3+\delta_8+\frac{1}{2}\delta_{13}\right)^r\pi_2^2+(\delta_4+\delta_9)^r\pi_3+(\delta_5+\delta_{10})^r\pi_4\right)\\
&\quad -2\pi_5\left((\delta_1-\delta_6)^i\pi_1^2+\left(\delta_2-\delta_7-\frac{1}{2}\delta_{12}\right)^i\pi_1\pi_2\right.\\
&\quad \left.+\left(\delta_3-\delta_8-\frac{1}{2}\delta_{13}\right)^i\pi_2^2+(\delta_4-\delta_9)^i\pi_3-\delta_{11}^r\pi_5\right)\\
&\quad -2\pi_4\pi_5(\delta_5-\delta_{10}+\delta_{11})^i+\pi_1\left(\frac{1}{2}\delta_{14}^r\pi_1^2\pi_2-\left(\frac{1}{4}\delta_{11}-\frac{1}{2}\delta_{15}\right)^r\pi_1\pi_2^2\right.\\
&\quad \left.+2\delta_{14}^r\pi_1\pi_3+(\delta_{11}+2\delta_{15})^r\pi_2\pi_3\right)\\
\mathrm{RFOT5} &= 2\pi_4\left((\delta_1-\delta_6)^i\pi_1^2+\left(\delta_2-\delta_7-\frac{1}{2}\delta_{12}\right)^i\pi_1\pi_2+\left(\delta_3-\delta_8-\frac{1}{2}\delta_{13}\right)^i\pi_2^2\right.\\
&\quad \left.+(\delta_4-\delta_9)^i\pi_3+(\delta_5-\delta_{10})^i\pi_4\right)\\
&\quad +2\pi_5\left((\delta_1+\delta_6)^r\pi_1^2+\left(\delta_2+\delta_7+\frac{1}{2}\delta_{12}\right)^r\pi_1\pi_2\right.\\
&\quad \left.+\left(\delta_3+\delta_8+\frac{1}{2}\delta_{13}\right)^r\pi_2^2+(\delta_4+\delta_9)^r\pi_3-\delta_{11}^i\pi_5\right)\\
&\quad +2\pi_4\pi_5(\delta_5+\delta_{10}-\delta_{11})^r+\pi_1\left(-\frac{1}{2}\delta_{14}^i\pi_1^2\pi_2+\left(\frac{1}{4}\delta_{11}-\frac{1}{2}\delta_{15}\right)^i\pi_1\pi_2^2\right.\\
&\quad \left.+2\delta_{14}^i\pi_1\pi_3+(\delta_{11}+2\delta_{15})^i\pi_2\pi_3\right).
\end{aligned}$$

Lemma 3.6.1 *Restriction to the stratum $\Delta = 0$ yields*

$$
\begin{aligned}
\Delta\mathrm{RFOT1} &= 2\pi_1\left(\delta_1^r\pi_1^2 + \delta_2^r\pi_1\pi_2 + \left(\delta_3 + \frac{1}{4}\delta_4\right)^r\pi_2^2\right)\\
&\quad +\pi_4\Big(2(\delta_5+\delta_{14})^r\pi_1 + (\delta_{11}+2\delta_{15})^r\pi_2\Big)\\
&\quad +\pi_5\Big(2\delta_{14}^i\pi_1 + (\delta_{11}+2\delta_{15})^i\pi_2\Big)\\
\Delta\mathrm{RFOT2} &= 2\pi_2\left(\delta_6^r\pi_1^2 + \left(\delta_7+\frac{1}{2}\delta_{12}\right)^r\pi_1\pi_2 + \left(\delta_8+\frac{1}{4}\delta_9+\frac{1}{2}\delta_{13}\right)^r\pi_2^2\right)\\
&\quad +\pi_4\Big(2\delta_{14}^r\pi_1 + (2\delta_{10}-\delta_{11}+2\delta_{15})^r\pi_2\Big)\\
&\quad -\pi_5\Big(2\delta_{14}^i\pi_1 + (\delta_{11}+2\delta_{15})^i\pi_2\Big)\\
\Delta\mathrm{RFOT3} &= \frac{1}{2}\pi_2\,\Delta\mathrm{RFOT2}\\
\Delta\mathrm{RFOT4} &= 2\pi_4\Bigg((\delta_1+\delta_6)^r\pi_1^2 + \left(\delta_2+\delta_7+\frac{1}{2}\delta_{12}\right)^r\pi_1\pi_2\\
&\quad +\left(\delta_3+\frac{1}{4}\delta_4+\delta_8+\frac{1}{4}\delta_9+\frac{1}{2}\delta_{13}\right)^r\pi_2^2 + (\delta_5+\delta_{10})^r\pi_4\Bigg)\\
&\quad -2\pi_5\Bigg((\delta_1-\delta_6)^i\pi_1^2 + \left(\delta_2-\delta_7-\frac{1}{2}\delta_{12}\right)^i\pi_1\pi_2\\
&\quad +\left(\delta_3+\frac{1}{4}\delta_4-\delta_8-\frac{1}{4}\delta_9-\frac{1}{2}\delta_{13}\right)^i\pi_2^2 - \delta_{11}^r\pi_5\Bigg)\\
&\quad -2\pi_4\pi_5(\delta_5-\delta_{10}+\delta_{11})^i + \pi_1\pi_2\Bigg(\frac{1}{2}\delta_{14}^r\pi_1^2\\
&\quad -\left(\frac{1}{4}\delta_{11}-\frac{1}{2}\delta_{14}-\frac{1}{2}\delta_{15}\right)^r\pi_1\pi_2 + \left(\frac{1}{4}\delta_{11}+\frac{1}{2}\delta_{15}\right)^r\pi_2^2\Bigg)\\
\Delta\mathrm{RFOT5} &= 2\pi_4\Bigg((\delta_1-\delta_6)^i\pi_1^2 + \left(\delta_2-\delta_7-\frac{1}{2}\delta_{12}\right)^i\pi_1\pi_2\\
&\quad +\left(\delta_3+\frac{1}{4}\delta_4-\delta_8-\frac{1}{4}\delta_9-\frac{1}{2}\delta_{13}\right)^i\pi_2^2 + (\delta_5-\delta_{10})^i\pi_4\Bigg)\\
&\quad +2\pi_5\Bigg((\delta_1+\delta_6)^r\pi_1^2 + \left(\delta_2+\delta_7+\frac{1}{2}\delta_{12}\right)^r\pi_1\pi_2\\
&\quad +\left(\delta_3+\frac{1}{4}\delta_4+\delta_8+\frac{1}{4}\delta_9+\frac{1}{2}\delta_{13}\right)^r\pi_2^2 - \delta_{11}^i\pi_5\Bigg)\\
&\quad +2\pi_4\pi_5(\delta_5+\delta_{10}-\delta_{11})^r + \pi_1\pi_2\Bigg(-\frac{1}{2}\delta_{14}^i\pi_1^2\\
&\quad +\left(\frac{1}{4}\delta_{11}+\frac{1}{2}\delta_{14}-\frac{1}{2}\delta_{15}\right)^i\pi_1\pi_2 + \left(\frac{1}{4}\delta_{11}+\frac{1}{2}\delta_{15}\right)^i\pi_2^2\Bigg).
\end{aligned}
$$

Here ΔRFOT1, ... , ΔRFOT5 denote the components of the reduced fifth order terms (RFOT) restricted to the statum $\Delta = 0$.

3.7 Singular perturbation theory

For the moment we want to restrict our considerations to the stratum $\Delta = 0$. The curve $g(\pi_1)$, $0 \leq \pi_1 \leq -\frac{\lambda}{\alpha^r}$, of critical points of the reduced vector field $\dot{\pi} = f(\pi)$ (cf. Equations 3.4.3 to 3.4.7) is located on this stratum (cf. Lemma 3.4.6). According to Lemma 3.5.2 this curve is asymptotically stable for our choice of the coefficients

$$\beta^r < \alpha^r < \gamma^r < 0.$$

Now we want to study the perturbed vector field (cf. Lemma 3.6.1)

$$\dot{\pi} \quad = \quad f^\varepsilon(\pi) \quad = \quad f(\pi) + \varepsilon\, \Delta\mathrm{RFOT}(\pi), \;\; \varepsilon \ll 1. \tag{3.7.19}$$

We have the following propostion.

Proposition 3.7.1 *For the perturbed Vector Field 3.7.19 and $0 < |\varepsilon| < \varepsilon_0$ there persists an invariant curve g_ε near g on the stratum $\Delta = 0$. This curve g_ε is parametrised over π_1. The vector field on g_ε has the form*

$$r(\pi_1) \quad = \quad 2\frac{\pi_1\pi_2}{\pi_1+\pi_2}\left(16\pi_1^2 + 16\frac{\lambda}{\alpha^r}\pi_1 + 3\frac{\lambda^2}{\alpha^{r2}}\right) d$$

with

$$0 < \pi_1 < -\frac{\lambda}{\alpha^r}, \;\; \pi_1 + \pi_2 = -\frac{\lambda}{\alpha^r},$$

$$d = \left((d_6 + d_7 - d_8)^r + \frac{(\alpha-\beta)^i}{(\alpha-\beta)^r}(d_6+d_7-d_8)^i\right).$$

Proof: In Lemma 3.5.2 we showed that the curve $g(\pi_1)$, $0 < \pi_1 < -\frac{\lambda}{\alpha^r}$, is normally hyperbolic. Thus an invariant curve g_ε near g persists under small perturbations. The curve g_ε will no longer consist of critical points but there will be a resulting flow on g_ε. This flow is determined in the lowest order by projection of the perturbation onto the curve g.
Let

$$E \quad = \quad \left\{ g(\pi_1) \,|\, 0 < \pi_1 < -\frac{\lambda}{\alpha^r} \right\}$$

be the curve of critical points of the vector field $f^0(\pi)$ on the stratum

$$F = \left\{ \pi \in \Pi(\mathbb{R}^6) \,|\, \Delta(\pi) = 0 \right\}.$$

For a point $\pi \in E$ let

$$Tf^0(\pi) \quad : \quad T_\pi F \;\; \to \;\; T_\pi F$$

denote the linearization of f^0 in π. By construcion $T_\pi E$ lies in the kernel of $Tf^0(\pi)$. So a linear map

$$Qf^0(\pi) \quad : \quad \frac{T_\pi F}{T_\pi E} \;\; \to \;\; \frac{T_\pi F}{T_\pi E}$$

is induced on the quotient space. The eigenvalues of $Qf^0(\pi)$ have been determined in Lemma 3.5.2 and are both negative. Thus for every $\pi \in E$ $T_\pi E$ has a unique complement N_π that is invariant under $Tf^0(\pi)$. Let P^E denote the projection onto TE defined by the splitting

$$TF_{|E} = TE \oplus N$$

On E we define the vector field

$$f_R(\pi) = P^E \frac{\partial}{\partial \varepsilon} f^\varepsilon(\pi)_{|\varepsilon=0}.$$

We have the extended vector field

$$f^\varepsilon(\pi) \times \{0\} \text{ on } F \times (-\varepsilon_0, \varepsilon_0).$$

In this system, according to [Fe], a two dimensional center manifold C exists for small $\varepsilon_0 \ll 1$. The second dimension has its origin in the extension of the system in ε-direction.
On C near $E \times \{0\}$ a smooth vector field

$$f_C = \begin{cases} \frac{1}{\varepsilon} f^\varepsilon(\pi) \times 0, & \varepsilon \neq 0 \\ f_R(\pi) \times 0, & \varepsilon = 0 \end{cases}$$

is defined. The center manifold is fibered in ε-direction with invariant curves g_ε. The flow on g_ε has the form

$$\dot{\pi} = \varepsilon f_R(\pi) + O(\varepsilon^2).$$

We want to determine the vector field

$$f_R(\pi) = P^E \frac{\partial}{\partial \varepsilon} f^\varepsilon(\pi)_{|\varepsilon=0}.$$

The vectors t_1, t_2, t_3 that span the tangent space to the stratum F along the curve g, are known (cf. Lemma 3.5.2). The vector t_1 is the tangent vector along the curve g. Now we want to write the terms of higher order ΔRFOT along the curve g in the form

$$\Delta\text{RFOT}(\pi) = a(\pi)t_1 + b(\pi)t_2 + c(\pi)t_3.$$

This gives the projection P^E we are looking for and we have

$$f_R(\pi) = a(\pi).$$

As the restriction of the vector field $\frac{\partial}{\partial\varepsilon}f^\varepsilon(\pi)_{|\varepsilon=0}$ along the curve g one gets

$$
\begin{aligned}
r_1 &= \pi_1\left(2\delta_1^r\pi_1^2 + (2\delta_2+\delta_5+\delta_{14})^r\pi_1\pi_2 + \left(2\delta_3+\frac{1}{2}\delta_4+\frac{1}{2}\delta_{11}+\delta_{15}\right)^r\pi_2^2\right)\\
r_2 &= \pi_2\Bigg((2\delta_6+\delta_{14})^r\pi_1^2 + \left(2\delta_7+\delta_{10}-\frac{1}{2}\delta_{11}+\delta_{12}+\delta_{15}\right)^r\pi_1\pi_2\\
&\quad+\left(2\delta_8+\frac{1}{2}\delta_9+\delta_{13}\right)^r\pi_2^2\Bigg)\\
r_3 &= \frac{1}{2}\pi_2\, r_2\\
r_4 &= \pi_1\pi_2\Bigg(\left(\delta_1+\delta_6+\frac{1}{2}\delta_{14}\right)^r\pi_1^2\\
&\quad+\left(\delta_2+\frac{1}{2}\delta_5+\delta_7+\frac{1}{2}\delta_{10}-\frac{1}{4}\delta_{11}+\frac{1}{2}\delta_{12}+\frac{1}{2}\delta_{14}+\frac{1}{2}\delta_{15}\right)^r\pi_1\pi_2\\
&\quad+\left(\delta_3+\frac{1}{4}\delta_4+\delta_8+\frac{1}{4}\delta_9+\frac{1}{4}\delta_{11}+\frac{1}{2}\delta_{13}+\frac{1}{2}\delta_{15}\right)^r\pi_2^2\Bigg)\\
r_5 &= \pi_1\pi_2\Bigg(\left(\delta_1-\delta_6-\frac{1}{2}\delta_{14}\right)^i\pi_1^2\\
&\quad+\left(\delta_2+\frac{1}{2}\delta_5-\delta_7-\frac{1}{2}\delta_{10}+\frac{1}{4}\delta_{11}-\frac{1}{2}\delta_{12}+\frac{1}{2}\delta_{14}-\frac{1}{2}\delta_{15}\right)^i\pi_1\pi_2\\
&\quad+\left(\delta_3+\frac{1}{4}\delta_4-\delta_8-\frac{1}{4}\delta_9+\frac{1}{4}\delta_{11}-\frac{1}{2}\delta_{13}+\frac{1}{2}\delta_{15}\right)^i\pi_2^2\Bigg).
\end{aligned}
$$

We always have

$$\pi_1+\pi_2 = -\frac{\lambda}{\alpha^r}$$

and get the following equations

$$
\begin{aligned}
r_1 &= a + b\pi_1 + 2c\alpha^r(\alpha-\beta)^i && (3.7.20)\\
r_2 &= -a + b\pi_2 - 2c\alpha^r(\alpha-\beta)^i && (3.7.21)\\
r_4 &= \frac{1}{2}a(\pi_2-\pi_1) + b\pi_1\pi_2 + c\alpha^r(\pi_2-\pi_1)(\alpha-\beta)^i && (3.7.22)\\
r_5 &= -c\alpha^r(\pi_1+\pi_2)(\alpha-\beta)^r. && (3.7.23)
\end{aligned}
$$

Thus

$$c = -\frac{r_5}{\alpha^r(\pi_1+\pi_2)(\alpha-\beta)^r},$$

$$r_1+r_2 = b(\pi_1+\pi_2)$$

and

$$r_1-r_2 = 2a + b(\pi_1-\pi_2) + 4c\alpha^r(\alpha-\beta)^i.$$

Finally we get

$$b = \frac{r_1 + r_2}{\pi_1 + \pi_2}$$

and

$$\begin{aligned} a &= \frac{1}{2}(r_1 - r_2) - \frac{1}{2}b(\pi_1 - \pi_2) - 2c\alpha^r(\alpha - \beta)^i \\ &= \frac{r_1\pi_2 - r_2\pi_1}{\pi_1 + \pi_2} - 2c\alpha^r(\alpha - \beta)^i. \end{aligned}$$

Insertion of r_1, r_2, c and retranslation of the coefficients $\delta_1, \ldots, \delta_{15}$ into the coefficients $d_1, \ldots, d_9$ finishes the proof. ⋈

In the following let

$$d \neq 0.$$

PROPOSITION 3.7.2 *On the invariant curve g_ε (cf. Propostion 3.7.1) for the perturbed Vector Field 3.7.19 exactly two critical points persist for $0 < |\varepsilon| < \tilde{\varepsilon}_0 < \varepsilon_0$. In the entire ten dimensional system these critical points have isotropy $(\mathrm{O}(2), 1)$ resp. $(\mathrm{D}_4, \mathbb{Z}_2)$. The latter corresponds to the isotropy $(\mathrm{D}_2, \mathbb{Z}_2)$ in the reduced system. Their stability in $\mathbb{R}^6$ is determined by the sign of*

$$d = \left((d_6 + d_7 - d_8)^r + \frac{(\alpha - \beta)^i}{(\alpha - \beta)^r}(d_6 + d_7 - d_8)^i\right).$$

Especially a connection between the group orbits of solutions with isotropy $(\mathrm{O}(2), 1)$ resp. $(\mathrm{D}_4, \mathbb{Z}_2)$ persists for small ε in $\mathbb{R}^6$.
The position of the critical points, their isotropy in the entire system, and the direction of the resulting flow on g_ε is given in Figure 5.

Proof: On the curve $g_\varepsilon(\pi_1), 0 < \pi_1 < -\frac{\lambda}{\alpha^r}$, near g there are two critical points of the Fenichel vector field $r(\pi_1)$ (cf. Proposition 3.7.1) with

$$\pi_1 \in \left\{-\frac{\lambda}{4\alpha^r}, -\frac{3\lambda}{4\alpha^r}\right\}.$$

Linearization of the vector field $r(\pi_1)$ in these critical points yields

π_1	$\frac{\mathrm{d}r}{\mathrm{d}\pi_1}$
$-\frac{\lambda}{4\alpha^r}$	$-3d\frac{\lambda^2}{\alpha^{r2}}$
$-\frac{3\lambda}{4\alpha^r}$	$3d\frac{\lambda^2}{\alpha^{r2}}$,

and, thus, they are hyperbolic. Here d is defined as in Proposition 3.7.1. Therefore these critical points persist for $|\varepsilon| < \tilde{\varepsilon}_0 < \varepsilon_0$ in the perturbed Vector Field 3.7.19.
We shall show that the persisting critical points lie on the group orbits of solutions

with isotropy $(\mathrm{O}(2),1)$ resp. $(\mathrm{D}_4,\mathbb{Z}_2)$ with reference to the entire system.
First let

$$\tilde{\pi} \quad = \quad -\frac{\lambda}{4\alpha^r}.$$

Using the representation of the group element

$$\tau \quad = \quad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

introduced in Lemma 3.3.3 and

$$\pi \quad \in \quad \mathrm{S}^1,$$

we have

$$\tau\,\pi\,(0,z,0) \quad = \quad \tau\,(0,-z,0) \;=\; \left(\frac{1}{2}\sqrt{\frac{3}{2}}z, \frac{1}{2}z, \frac{1}{2}\sqrt{\frac{3}{2}}z\right).$$

Points of the form $(0,z,0)$ with

$$|z|^2 \quad = \quad -\frac{\lambda}{\alpha^r}$$

are mapped to the critical point of isotropy $(\mathrm{O}(2),1)$ in the reduced system by the Hilbert-map. Thus

$$\begin{aligned} \Pi(\tau\pi\,(0,z,0)) \quad = \quad & \Bigg(\pi_1 = -\frac{\lambda}{4\alpha^r}, \pi_2 = -\left(\pi_1 + \frac{\lambda}{\alpha^r}\right), \pi_3 = \frac{1}{4}\pi_2^2, \\ & \pi_4 = \frac{1}{2}\pi_1\pi_2, \pi_5 = 0\Bigg). \end{aligned}$$

Therefore

$$\Pi\big(\tau\,\pi\,(0,z,0)\big) \quad = \quad g(\tilde{\pi}).$$

Second let

$$\tilde{\pi} \quad = \quad -\frac{3\lambda}{4\alpha^r}.$$

Correspondingly the Hilbert-map maps points of the form $(z,0,z)$ with

$$|z|^2 \quad = \quad -\frac{\lambda}{2\alpha^r}$$

to the critical point of isotropy $(\mathrm{D}_2,\mathbb{Z}_2)$ in the reduced system. With $r_\pi \in \mathrm{O}(2)$ we have

$$r_\pi\,\tau\,(z,0,z) \quad = \quad r_\pi\left(-\frac{1}{2}z, \sqrt{\frac{3}{2}}z, -\frac{1}{2}z\right) \;=\; \left(\frac{1}{2}z, \sqrt{\frac{3}{2}}z, \frac{1}{2}z\right).$$

Thus

$$\Pi(r_\pi\,\tau\,(z,0,z)) \;=\; \left(\pi_1 = -\frac{3\lambda}{4\alpha^r}, \pi_2 = -\left(\pi_1 + \frac{\lambda}{\alpha^r}\right), \pi_3 = \frac{1}{4}\pi_2^2, \right.$$
$$\left. \pi_4 = \frac{1}{2}\pi_1\pi_2, \pi_5 = 0\right) \;=\; g(\tilde{\pi}).$$

Since the perturbation respects the symmetry, the critical points persisting for small ε on the curve have the same isotropies.
Besides these two critical points there are no critical points on g_ε for small $\varepsilon > 0$. Since the two critical points are hyperbolic, in a neighbourhood of these points no further critical points exist by the implicit function theorem. If there were critical points (x_n, ε_n) in the remaining part of g_ε, for a sequence $(\varepsilon_n) \to 0$ the accumulation point $(\bar{x}, 0)$ would have to be a critical point of the resulting vector field in contradiction to Proposition 3.7.1. ⋈

Figure 5 shows the resulting flow on the invariant curve $g_\varepsilon(\pi_1)$, $0 < \pi_1 < -\frac{\lambda}{\alpha^r}$, in a schematic way for $d > 0$ and small $\varepsilon > 0$. Choosing $d < 0$ will change the direction of the arrows. The isotropies of the solutions in the entire ten dimensional system are indicated in the sketch.
For $\varepsilon = 0$ (i.e. $g_\varepsilon = g$) $g(0)$ resp. $g(-\frac{\lambda}{\alpha^r})$ are fixed points of isotropy $(\mathrm{D}_2, \mathbb{Z}_2)$ resp. $(\mathrm{O}(2), 1)$ (cf. Lemma 3.4.6). The curve itself consists of fixed points.

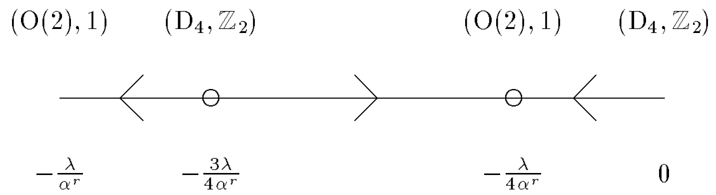


Figure 5: Resulting flow on g_ε

3.8 INVARIANT TORI

In this section we want to show that for small $\varepsilon > 0$ a fixed point bifurcates from the critical point $\tilde{\pi}_1$ in the direction of the principal stratum. The critical point $\tilde{\pi}_1$ lies on the curve g on the stratum $\Delta = 0$. According to Lemma 3.5.3 the linearization of the vector field of the reduced equation (cf. Lemma 3.2.1) has a nontrivial Jordan-block to the eigenvalue zero in the point

$$g(\tilde{\pi}_1), \;\; \tilde{\pi}_1 = -\frac{\lambda(\alpha-\gamma)^r}{\alpha^r(\beta-\gamma)^r}.$$

The position of this point on the invariant curve

$$g(\pi_1), \;\; 0 < \pi_1 < -\frac{\lambda}{\alpha^r},$$

depends on the relative choice of the coefficients

$$\beta^r < \alpha^r < \gamma^r < 0, \quad \alpha^r \in \left(\frac{1}{4}(\beta^r + 3\gamma^r), \gamma^r \right),$$

according to Figure 6. Making the ansatz

$$\alpha^r \;=\; t\,\beta^r + (1-t)\,\gamma^r, \quad t \in \left(0, \frac{1}{4}\right),$$

this follows as in the proof of Lemma 3.5.1. Thus

$$\tilde{\pi}_1 = -\frac{\lambda(\alpha-\gamma)^r}{\alpha^r(\beta-\gamma)^r} = -\frac{\lambda}{\alpha^r}t, \quad t \in \left(0, \frac{1}{4}\right).$$

Figure 6: Possible region of the point $\tilde{\pi}_1$

We want to determine the form of the resulting vector field on the local two dimensional center manifold W^c_{loc} near the point $g(\tilde{\pi}_1)$. The center manifold W^c_{loc} is tangential to Span$\,(t_1, h)$ (cf. Lemma 3.5.3) and intersects the stratum $\Delta = 0$ in a part of the invariant curve $g(\pi_1)$ near $g(\tilde{\pi}_1)$. Let t_1 be the tangent vector in the direction of the curve $g(\pi_1)$ and h be the hauptvector associated to the Jordan-block of the linearization. By definition of the vectors t_1, t, h in Lemma 3.5.3 h points in the direction of the principal stratum.
We introduce x-coordinates in the direction of $(-t_1)$ along the invariant curve $g(\pi_1)$ and y-coordinates in the direction of $(-h)$ with origin in $g(\tilde{\pi}_1)$. Therefore the vector field on W^c_{loc} has the form

$$\begin{aligned} \dot{x} &= -y + H(x, y) \\ \dot{y} &= y\, G(x, y). \end{aligned} \tag{3.8.24}$$

We are only interested in the region $y \leq 0$ that describes a part of the Hilbert-set $\Pi(\mathbb{R}^6)$ according to our choice of the coordinates. The $(-y)$-term in the x-equation models the Jordan-block, the minus sign follows from the equation

$$Lh \;=\; j\,t_1$$

with $j < 0$ according to Lemma 3.5.3. The y-term in the y-equation describes the flow invariance of the curve $y = 0$, i.e. of the stratum $\Delta = 0$.

The function $H(x,y)$ has the following properties

$$\begin{aligned} H(x,y) &= O(x^2, xy, y^2), \\ H(x,0) &\equiv 0, \\ \frac{\partial H}{\partial x}(x,0) &\equiv 0. \end{aligned}$$

The last two properties are due to the fact that points of the form $(x,0)$ are critical points of the system 3.8.24 by construction. The linearization of the Vector Field 3.8.24 in such a point $(x,0)$ yields

$$A = \begin{pmatrix} 0 & -1 + \frac{\partial H}{\partial y}(x,0) \\ 0 & G(x,0) \end{pmatrix}.$$

Consequently the eigenvalues are zero in the direction of the curve of fixed points and $G(x,0)$ in the direction of the principal stratum. This eigenvalue has been calculated in Lemma 3.5.3, and has the form

$$e_4 = 4\frac{\lambda(\alpha-\gamma)^r + \pi_1 \alpha^r (\beta-\gamma)^r}{\alpha^r}.$$

Therefore in our coordinates we have

$$G(x,0) = a\,x + O(x^2)$$

with $a > 0$. The invariant curve changes the stability in the direction of the principal stratum in the first order from stable to unstable in the point $(0,0)$ (transversality condition).

Now let's look at the extended system

$$\begin{aligned} \dot{\pi} &= f(\pi) + \varepsilon\,\mathrm{RFOT}(\pi) \\ \dot{\varepsilon} &= 0. \end{aligned} \tag{3.8.25}$$

Here near the point

$$(g(\tilde{\pi}_1), 0)$$

there exists a local center manifold. This manifold is fibered in ε-direction with two dimensional invariant manifolds $W^c_{loc,\varepsilon}$. For $\varepsilon = 0$ the manifold $W^c_{loc,0}$, tangential to Span$\{t_1, h\}$, intersects the stratum $\Delta = 0$ in a part of the curve g near $g(\tilde{\pi}_1)$ transversally. This property is preserved for small

$$\varepsilon < \overline{\varepsilon} < \tilde{\varepsilon}.$$

On the two dimensional center manifolds $W^c_{loc,\varepsilon}$ again we introduce, now ε-dependent, coordinates x_ε in the direction of g_ε and y_ε in the direction of the principal stratum. We shall continue writing x resp. y for x_ε resp. y_ε.

Now let $\tilde{\pi}_1$ be tuned in such a way such that the Fenichel-drift in $g_\varepsilon(\tilde{\pi}_1)$, $|\varepsilon| < \overline{\varepsilon}$, is not zero. Then the flow on the corresponding center manifold $W^c_{loc,\varepsilon}$ has the form

$$\begin{aligned} \dot{x} &= -y + \varepsilon + H(x,y,\varepsilon) \\ \dot{y} &= y\,G(x,y,\varepsilon). \end{aligned} \tag{3.8.26}$$

The sign of ε depends on the direction of the resulting Fenichel-drift. We want to assume the solution of isotropy $(\mathrm{O}(2),1)$ to be stable. Therefore according to Proposition 3.7.2 we have to choose $d > 0$ and the resulting Fenichel-drift has the form indicated in Figure 5. For the choice of parameters

$$\beta^r < \alpha^r < \gamma^r < 0, \quad \alpha^r \in \left(\frac{1}{4}(\beta+3\gamma)^r, \gamma^r\right),$$

we have

$$\tilde{\pi}_1 \quad \in \quad \left(0, -\frac{\lambda}{4\alpha^r}\right)$$

and, thus, we have to choose $\varepsilon < 0$.
The functions $G(x,y,\varepsilon)$ resp. $H(x,y,\varepsilon)$ have the following properties

$$\begin{aligned} G(x,y,\varepsilon) &= O(x,y,\varepsilon), \\ G(x,y,0) &= G(x,y) \end{aligned}$$

resp.

$$\begin{aligned} H(x,y,\varepsilon) &= O(x^2,xy,y^2,\varepsilon x,\varepsilon y,\varepsilon^2), \\ H(x,y,0) &= H(x,y). \end{aligned}$$

Proposition 3.8.1 *Let*

$$\beta^r < \alpha^r < \gamma^r < 0, \quad \alpha^r \in \left(\frac{1}{4}(\beta+3\gamma)^r, \gamma^r\right).$$

Then there exists $\overline{\varepsilon} > 0$ and a unique curve

$$\Big(x(\varepsilon), y(\varepsilon) \le 0\Big), \quad -\overline{\varepsilon} < \varepsilon \le 0,$$

of critical points of the flow on the center manifold $W^c_{loc,\varepsilon} \cap \Pi(\mathbb{R}^6)$ with

$$\Big(x(0), y(0)\Big) = (0,0).$$

The critical points are saddles.

Proof: We are looking for critical points of the Vector Field 3.8.26. Therefore we first solve the equation

$$P(x,y,\varepsilon) \quad = \quad -y+\varepsilon+H(x,y,\varepsilon) \;=\; 0.$$

We have

$$P(0,0,0) \quad = \quad 0$$

and

$$\frac{\partial P}{\partial y}(0,0,0) \quad = \quad -1$$

since $H(x,y,\varepsilon)$ is of second order. Using the implicit function theorem, locally near $(x,\varepsilon) = (0,0)$ one gets a unique surface $y = y(x,\varepsilon)$ with $y(0,0) = 0$ and $P(x, y(x,\varepsilon), \varepsilon) = 0$. Furthermore

$$y(x,\varepsilon) \;=\; \varepsilon + O(x^2, x\varepsilon, \varepsilon^2)$$

and

$$\frac{\partial y}{\partial x}(0,0) \;=\; 0.$$

Now we want to solve the equation

$$G(x, y(x,\varepsilon), \varepsilon) \;=\; 0.$$

We have

$$G(0,0,0) \;=\; 0$$

and

$$\frac{\partial G}{\partial x}(0,0,0) \;=\; a > 0$$

because of the transversality property of G and the condition $\frac{\partial y}{\partial x}(0,0) = 0$. Therefore, again by the implicit function theorem, there exists a unique curve

$$(x(\varepsilon), y(\varepsilon)),\;\; 0 \leq |\varepsilon| < \overline{\varepsilon},\;\; \varepsilon \leq 0,$$

of critical points of the Vector Field 3.8.26. Furthermore

$$x \;=\; O(\varepsilon).$$

Thus the curve $y(\varepsilon)$ has the form

$$y(\varepsilon) \;=\; \varepsilon + O(\varepsilon^2).$$

The sign of $y(\varepsilon)$ is determined by the sign of ε for small ε. Here we have $\varepsilon < 0$ and, therefore, $y(\varepsilon) < 0$. Consequently the curve lies in the Hilbert-set $\Pi(\mathbb{R}^6)$.
The linear stability of the critical point $(x(\varepsilon), y(\varepsilon))$, $-\overline{\varepsilon} < \varepsilon \leq 0$, is to be determined. The linearization of the Vector Field 3.8.26 in the point $(x(\varepsilon), y(\varepsilon))$ yields

$$\begin{aligned} D \;&=\; \begin{pmatrix} \frac{\partial H}{\partial x}(x(\varepsilon), y(\varepsilon), \varepsilon) & -1 + \frac{\partial H}{\partial y}(x(\varepsilon), y(\varepsilon), \varepsilon) \\ y(\varepsilon)\, \frac{\partial G}{\partial x}(x(\varepsilon), y(\varepsilon), \varepsilon) & G(x(\varepsilon), y(\varepsilon), \varepsilon) + y(\varepsilon)\, \frac{\partial G}{\partial y}(x(\varepsilon), y(\varepsilon), \varepsilon) \end{pmatrix} \\ &=\; \begin{pmatrix} O(\varepsilon) & -1 + O(\varepsilon) \\ \varepsilon\, a(\varepsilon) + O(\varepsilon^2) & O(\varepsilon) \end{pmatrix}. \end{aligned}$$

We have $\frac{\partial G}{\partial x}(0,0,0) = a > 0$. Thus

$$\frac{\partial G}{\partial x}(x(\varepsilon), y(\varepsilon), \varepsilon) \;=\; a(\varepsilon) \;>\; 0$$

with $a(0) = a$ for small ε. So we get two eigenvalues of D of the following form

$$\rho_{1,2} = O(\varepsilon) \pm \sqrt{O(\varepsilon^2) - \varepsilon\, a(\varepsilon)}$$

with $\varepsilon < 0$ and $a(\varepsilon) > 0$. For small ε the $\sqrt{-\varepsilon}$-term is dominating, the critical point is a saddle. $\bowtie$

The bifurcating critical point lies in the principal stratum. The preimages are two 2-tori. Since there are no additional, symmetry given phase relations (cf. Lemma 3.4.9) in general we have quasiperiodic solutions.

3.9 STABILITY OF THE INVARIANT TORI

We want to know the stability of the group orbit of the quasiperiodic solutions (cf. Proposition 3.8.1) in the entire ten dimensional system. This information is useful for calculating the Conley-index of this group orbit (cf. [Le]). We shall determine the Floquet-exponents of the periodic solutions that correspond to the critical points on the curve

$$g(\pi_1), \quad 0 < \pi_1 < -\frac{\lambda}{\alpha^r}.$$

According to our choice of the coefficients only the interval

$$0 < \pi_1 < -\frac{\lambda}{4\alpha^r}$$

is of interest. Here, in dependence on the relative choice of the coefficients, critical points of the reduced system bifurcate (cf. Proposition 3.8.1).

The periodic solutions are rotating waves. In a rotating coordinate system one gets a static problem which is accessible more easily. We make the ansatz

$$\begin{aligned} z_0 &= (r_0 + \rho_0)\, e^{\mathrm{i}(\omega_0\, t + \phi_0)} \\ z_{\pm 2} &= (r_2 + \rho_{\pm 2})\, e^{\mathrm{i}(\omega_0\, t + \phi_{\pm 2})} \\ z_{\pm 1} &= y_{\pm 1}\, e^{\mathrm{i}\omega_0\, t} \end{aligned}$$

with

$$\omega_0 = \omega + \alpha^i \left(r_0^2 + 2r_2^2\right)$$

and

$$r_0^2 + 2r_2^2 = -\frac{\lambda}{\alpha^r}.$$

In the lowest order one gets the following systems which decouple for symmetry reasons:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} y_1 \\ \overline{y}_1 \\ y_{-1} \\ \overline{y}_{-1} \end{pmatrix} = \begin{pmatrix} s & t & t & s \\ \overline{t} & \overline{s} & \overline{s} & \overline{t} \\ t & s & s & t \\ \overline{s} & \overline{t} & \overline{t} & \overline{s} \end{pmatrix} \begin{pmatrix} y_1 \\ \overline{y}_1 \\ y_{-1} \\ \overline{y}_{-1} \end{pmatrix}$$

with

$$\begin{aligned} s &= r_0^2\left(-\alpha+\frac{1}{4}\beta+\frac{3}{4}\gamma\right)+r_2^2\left(-2\alpha+\frac{3}{2}\beta+\frac{1}{2}\gamma\right), \\ t &= 2\sqrt{\frac{3}{8}}(\gamma-\beta)r_0r_2 \end{aligned}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\rho_{-2}\\ \rho_0\\ \rho_2\\ \phi_{-2}\\ \phi_0\\ \phi_2\end{pmatrix} = \begin{pmatrix} 2r_2^2\gamma^r-r_0^2(\alpha-\beta)^r & 2r_0r_2\alpha^r & -2r_2^2\gamma^r+r_0^2(\alpha-\beta)^r+4r_2^2\alpha^r & r_0^2r_2(\alpha-\beta)^i & -2r_0^2r_2(\alpha-\beta)^i & r_0^2r_2(\alpha-\beta)^i \\ 2r_0r_2\alpha^r & 2r_0^2\alpha^r & 2r_0r_2\alpha^r & -2r_0r_2^2(\alpha-\beta)^i & 4r_0r_2^2(\alpha-\beta)^i & -2r_0r_2^2(\alpha-\beta)^i \\ -2r_2^2\gamma^r+r_0^2(\alpha-\beta)^r+4r_2^2\alpha^r & 2r_0r_2\alpha^r & 2r_2^2\gamma^r-r_0^2(\alpha-\beta)^r & r_0^2r_2(\alpha-\beta)^i & -2r_0^2r_2(\alpha-\beta)^i & r_0^2r_2(\alpha-\beta)^i \\ 2r_2\gamma^i-\frac{r_0^2}{r_2}(\alpha-\beta)^i & 2r_0\alpha^i & -2r_2\gamma^i+\frac{r_0^2}{r_2}(\alpha-\beta)^i+4r_2\alpha^i & -(\alpha-\beta)^r r_0^2 & 2(\alpha-\beta)^r r_0^2 & -(\alpha-\beta)^r r_0^2 \\ 2r_2\alpha^i & 2r_0\alpha^i & 2r_2\alpha^i & 2(\alpha-\beta)^r r_2^2 & -4(\alpha-\beta)^r r_2^2 & 2(\alpha-\beta)^r r_2^2 \\ -2r_2\gamma^i+\frac{r_0^2}{r_2}(\alpha-\beta)^i+4r_2\alpha^i & 2r_0\alpha^i & 2r_2\gamma^i-\frac{r_0^2}{r_2}(\alpha-\beta)^i & -(\alpha-\beta)^r r_0^2 & 2(\alpha-\beta)^r r_0^2 & -(\alpha-\beta)^r r_0^2 \end{pmatrix}\begin{pmatrix}\rho_{-2}\\ \rho_0\\ \rho_2\\ \phi_{-2}\\ \phi_0\\ \phi_2\end{pmatrix}.$$

One gets the following eigenvalues

$$\begin{aligned} \mu_{1,2} &= 0, \\ \mu_3 &= 2(s+t)^r, \\ \mu_4 &= 2(s-t)^r. \end{aligned}$$

Our choice of coordinates yields

$$\begin{aligned} \left(-\alpha+\frac{1}{4}\beta+\frac{3}{4}\gamma\right)^r &< 0, \\ \left(-\alpha+\frac{3}{4}\beta+\frac{1}{4}\gamma\right)^r &< 0, \\ 2\sqrt{\frac{3}{8}}(\gamma-\beta)^r &> 0, \end{aligned}$$

and, therefore,

$$\mu_4 < 0.$$

Finally we want to show

$$\mu_3 \quad < \quad 0.$$

By insertion on gets

$$\mu_3 \quad = \quad 2\left((\gamma-\beta)^r\left(\frac{1}{2}r_0^2+2\sqrt{\frac{3}{8}}r_0\sqrt{-\frac{\lambda}{2\alpha^r}-\frac{r_0^2}{2}}\right)-\frac{\lambda}{2\alpha^r}\left(-2\alpha+\frac{3}{2}\beta+\frac{1}{2}\gamma\right)^r\right).$$

The ansatz

$$r_0^2 \quad = \quad -t\frac{\lambda}{\alpha^r}, \quad t\in\left(0,\frac{1}{4}\right),$$

yields

$$\mu_3 \quad = \quad -2\frac{\lambda}{\alpha^r}\left((\gamma-\beta)^r\left(\frac{1}{2}t+2\sqrt{\frac{3}{8}}\sqrt{\frac{t(1-t)}{2}}\right)+\frac{1}{2}\left(-2\alpha+\frac{3}{2}\beta+\frac{1}{2}\gamma\right)^r\right).$$

In the admissible region we have

$$\mu_3 \quad < \quad 0.$$

The eigenvalues of the second system are (cf. Lemmata 3.5.2 and 3.5.3),

$$\begin{aligned}
\mu_{1,2,3} &= 0,\\
\mu_4 &= -2\lambda<0,\\
\mu_5 &= 2\frac{\lambda}{\alpha^r}(\alpha-\beta)^r<0,\\
\mu_6 &= -2\left(r_0^2(\alpha-\beta)^r+2r_2^2(\alpha-\gamma)^r\right).
\end{aligned}$$

Therefore in the bifurcation point

$$\tilde{\pi}_1 \quad = \quad -\frac{\lambda(\alpha-\gamma)^r}{\alpha^r(\beta-\gamma)^r}$$

there are six trivial and four negative Floquet-exponents. In the entire system the solution has isotropy $(\mathrm{D}_2, 1)$ in the bifurcation point. Thus the group orbit is four dimensional. Therefore four trivial exponents are symmetry given. The sign of the Floquet-exponents of the periodic solution corresponds to the sign of the eigenvalues of the associated fixed point in the stratified space. Dealing with fixed point bifurcation in the stratified space the group orbit of the bifurcating solution inherits the stability of the bifurcation point. The double zero eigenvalue at the bifurcation point splits into one positive and one negative eigenvalue (cf. Proposition 3.8.1). Therefore the bifurcating fixed point is hyperbolic. In the entire system the bifurcating fixed point has isotropy $(\mathbb{Z}_2, 1)$.
The following lemma is shown.

Lemma 3.9.1 *The bifurcating group orbit (cf. Proposition 3.8.1) has the unstable dimension one.*

REFERENCES

[Bi] *Bierstone E.*, Lifting isotopies from orbit spaces, Topology 14 (1975), 245–252.

[BrtD] *Bröcker Th. and T. tom Dieck*, Representations of Compact Lie Groups, GTM, Springer, 1985.

[ChLaMe] *Chossat P., R. Lauterbach and I. Melbourne*, Steady-State Bifurcation with O(3)-Symmetry, Arch. Rat. Mech. Anal. 113 (1990), 313–376.

[Co] *Conley Ch.*, Isolated Invariant Sets and the Morse Index, Conf. Board Math. Sci. 38, AMS, 1978.

[Di] *Dieudonné J.*, Grundzüge der modernen Analysis 3, Vieweg, 1975.

[EletAl] *Elphick C., E. Tirapegui, M.E. Brachet, P. Coullet and G. Iooss*, A simple global characterisation for normal forms of singular vector fields, Physica D 29 (1987), 95–127.

[Fe] *Fenichel N.*, Geometric Singular Pertubation Theory for Ordinary Differential Equations, J. Diff. Eqns. 31 (1979), 53–89.

[Fi] *Fiedler B.*, Global Bifurcations of Periodic Solutions with Symmetry, Lecture Notes in Math. 1309, Springer, 1988.

[FiMi] *Fiedler B. and K. Mischaikow*, Dynamics of Bifurcations for Variational Problems with O(3) Equivariance: A Conley Index Approach, Arch. Rat. Mech. Anal. 119 (1992), 145–196.

[HaRoSt] *Haaf H., M. Roberts and I. Stewart*, A Hopf bifurcation with symmetry, ZAMP 43 (1992), 793–826.

[IoRo] *Iooss G. and M. Rossi*, Hopf bifurcation in the presence of spherical symmetry: Analytical Results, SIAM J. Math. Anal. 20 No 3 (1989), 511–532.

[GoSt] *Golubitsky M. and I. Stewart*, Hopf Bifurcation in the Presence of Symmetry, Arch. Rat. Mech. Anal. 87 No 2 (1985), 107–165.

[GoStSch] *Golubitsky M., I. Stewart and D. Schaeffer*, Singularities and Groups in Bifurcation Theory II, Appl. Math. Sci. 69, Springer, 1988.

[La1] *Lauterbach R.*, Bifurcation with O(3)-symmetry, Habilitationsschrift, Universität Augsburg, 1988.

[La2] *Lauterbach R.*, Äquivariante Dynamische Systeme, Skript, FU-Berlin, 1995.

[Le] *Leis Ch.*, Hopf-Bifurcation in Systems with Spherical Symmetry Part II : Connecting Orbits, to appear in J. reine angew. Math.

[Mi] *Miller W.*, Symmetry groups and their application, Acad. Press., 1972.

[MoRoSt] *Montaldi J., M. Roberts and I. Stewart*, Stability of nonlinear normal modes of symmetric Hamiltonian systems, Nonlinearity 3 (1990), 731–722.

[Sa] *Sattinger D.H.*, Group Theoretic Methods in Bifurcation Theory, Lecture Notes in Math. 762, Springer, 1979.

[StiFä] *Stiefel E. and A. Fässler*, Gruppentheoretische Methoden und ihre Anwendung, Teubner, 1979.

Christian Leis
FU-Berlin
Institut für Mathematik
Arnimallee 2-6
14195 Berlin, Germany
E-Mail: leismath.fu-berlin.de

Semigroup Crossed Products and Hecke Algebras Arising from Number Fields

Jane Arledge, Marcelo Laca and Iain Raeburn[1]

Received: July 15, 1996
Revised: May 27, 1997

Communicated by Joachim Cuntz

Abstract. Recently Bost and Connes considered a Hecke C^*-algebra arising from the ring inclusion of $\mathbb{Z}$ in $\mathbb{Q}$, and a C^*-dynamical system involving this algebra. Laca and Raeburn realized this algebra as a semigroup crossed product, and studied it using techniques they had previously developed for studying Toeplitz algebras. Here we associate Hecke algebras to general number fields, realize them as semigroup crossed products, and analyze their representations.

1991 Mathematics Subject Classification: Primary 46L55, Secondary 11R04, 22D25

Keywords: semigroup dynamical system, covariant representation, Hecke algebra

Introduction

In their work on phase transitions in number theory, Bost and Connes considered the Hecke algebra $\mathcal{H}(\Gamma, \Gamma_0)$ of a particular group–subgroup pair (Γ, Γ_0), and gave a presentation of this algebra involving a unitary representation of the additive group $\mathbb{Q}/\mathbb{Z}$ and an isometric representation of the multiplicative semigroup $\mathbb{N}^*$ [3]. From this presentation, Laca and Raeburn recognized $\mathcal{H}(\Gamma, \Gamma_0)$ as a dense subalgebra of a semigroup crossed product of the form $C^*(\mathbb{Q}/\mathbb{Z}) \rtimes \mathbb{N}^*$, and then applied techniques they had previously developed for studying Toeplitz algebras to obtain information about $\mathcal{H}(\Gamma, \Gamma_0)$ and its representations [8].

The fascinating ideas of Bost and Connes raise many possibilities for fruitful interaction between number theory and operator algebras, and in particular promise to provide new and intriguing examples of dynamical systems. Here we investigate a family of semigroup crossed products similar to $C^*(\mathbb{Q}/\mathbb{Z}) \rtimes \mathbb{N}^*$, but with $\mathbb{Q}$ replaced by a finite extension K of $\mathbb{Q}$, and the subring $\mathbb{Z}$ of $\mathbb{Q}$ replaced by the ring $\mathcal{O}$ of integers in K. We construct an action α of the multiplicative semigroup of nonzero integers $\mathcal{O}^\times$ on the C^*-algebra of the additive group $K/\mathcal{O}$, and show that all the main

[1]This research was supported by the Australian Research Council.

results of [8] carry over to an arbitrary number field K. This has not been completely routine: in particular, to construct some of the key representations and prove our main theorem we had to look very closely at the compact dual $(K/\mathcal{O})\widehat{}$ of the discrete Abelian group $K/\mathcal{O}$, and our results here may be of independent interest.

The main theorem of [8], motivated by our earlier approach to uniqueness theorems for semigroups of non-unitary isometries [1, 7], is a characterization of faithful representations of the crossed product $C^*(\mathbb{Q}/\mathbb{Z}) \rtimes \mathbb{N}^*$. Thus the crossed product has several faithful realizations: on $\ell^2(\mathbb{Q}/\mathbb{Z})$, extending the regular representation of $C^*(\mathbb{Q}/\mathbb{Z})$; on $\ell^2(\mathbb{N}^*)$, extending the Toeplitz representation of $\mathbb{N}^*$; and on $\ell^2(\Gamma_0\backslash\Gamma)$, arising from the canonical representation of $\mathcal{H}(\Gamma, \Gamma_0)$ in the commutant of the induced representation $\mathrm{Ind}_{\Gamma_0}^{\Gamma}1$. For our action α of $\mathcal{O}^\times$ by endomorphisms of $C^*(K/\mathcal{O})$, it is easy enough to construct the regular representation on $\ell^2(K/\mathcal{O})$. We shall find a group–subgroup pair $(\Gamma_K, \Gamma_\mathcal{O})$ whose Hecke algebra is isomorphic to our crossed product and hence gives a representation on $\ell^2(\Gamma_\mathcal{O}\backslash\Gamma_K)$, and, through our analysis of $(K/\mathcal{O})\widehat{}$, find faithful representations of $C^*(K/\mathcal{O})$ on $\ell^2(\mathcal{O}^\times)$ which are compatible with the Toeplitz representation of $\mathcal{O}^\times$. Our main theorem implies that all these realizations of $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ are faithful.

We begin in §1 by constructing the action α of $\mathcal{O}^\times$ on $C^*(K/\mathcal{O})$. For $a \in \mathcal{O}^\times$, α_a is determined on generators δ_y for $C^*(K/\mathcal{O})$ by averaging in the group algebra the generators δ_x corresponding to solutions of the equation $ax = y$ in $\mathcal{O}$; thus α is almost by definition a right inverse for the action of $\mathcal{O}^\times$ induced by multiplication on $K/\mathcal{O}$. We then discuss the crossed product $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$, which is universal for covariant representations of the system $(C^*(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$, and the dual action of $(K^*)\widehat{}$, which integrates to give a faithful expectation of $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ onto $C^*(K/\mathcal{O})$. We can immediately write down several representations of the crossed product, including the regular representation on $\ell^2(K/\mathcal{O})$.

In §2 we construct the Hecke algebra realization $\mathcal{H}(\Gamma_K, \Gamma_\mathcal{O})$ of the crossed product, and give a presentation of this algebra similar to that given by Bost and Connes in the case $K = \mathbb{Q}$. The isomorphism of $\mathcal{H}(\Gamma_K, \Gamma_\mathcal{O})$ into $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ gives a natural representation of the crossed product on $\ell^2(\Gamma_\mathcal{O}\backslash\Gamma_K)$, which we call the Hecke representation. It is interesting to note that, by identifying a subrepresentation with the GNS-representation of a faithful state on $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$, we can see directly that the Hecke representation is faithful. This approach bypasses the appeal to the theory of groupoid C^*-algebras in [3], and our own main theorem.

Our main technical innovations are in §3, where we discuss characters of $K/\mathcal{O}$. In [3] and [8], essential use was made of the injective character $r \mapsto \exp 2\pi i r$ on $\mathbb{Q}/\mathbb{Z}$. In general there are no injective characters, and one is forced to look for a family of characters which can play the same rôle. We show that there is a nonempty set $\mathcal{X}_K$ of characters χ with two important properties: $\chi(\mathfrak{a}^{-1}/\mathcal{O}) \neq 1$ for every nontrivial ideal $\mathfrak{a}$ in $\mathcal{O}$, and $\{r \mapsto \chi(br) : b \in \mathcal{O}\}$ is dense in $(K/\mathcal{O})\widehat{}$. The key step in the proof that $\mathcal{X}_K \neq \emptyset$ is the construction of projections which behave as one would expect $\alpha_\mathfrak{a}(1)$ to behave — if we knew that the action α extended to an action of the semigroup of ideals in $\mathcal{O}$. Using the characters in $\mathcal{X}_K$, we can construct representations of the crossed product on $\ell^2(\mathcal{O}^\times)$ extending the Toeplitz representation.

The characterization of faithful representations of $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ is Theorem 4.1. This theorem and its proof have a long history: the strategy is that used by Cuntz in [4], which has been streamlined over the years, and adapted to the present situation in [1, 7]. The crucial ingredient is an estimate, whose proof uses in several key places the properties of the characters in $\mathcal{X}_K$. Thus the end result is substantially deeper than its analogue in [8]; in addition, the presence of units in $\mathcal{O}^\times$, which is necessary for the construction of the action α, complicates the proof of the estimate. We finish §4 with a discussion of the various representations and their interrelations.

In our last section, we consider a field K with class number 1. Now the ring $\mathcal{O}$ is a principal ideal domain, and one can realize the semigroup of ideals in $\mathcal{O}$ as a subsemigroup S of $\mathcal{O}^\times$. There is therefore a similar dynamical system $(C^*(K/\mathcal{O}), S, \alpha)$ which does not involve units. The corresponding version of Theorem 4.1 is therefore slightly easier to prove, and is a direct generalization of the main theorem of [8].

While we were preparing the final version of this paper, we received a preprint from David Harari and Eric Leichtnam, in which they extend the original Bost-Connes analysis to more general fields K [5]. They associate a Hecke algebra to a class of fields more general than ours; however, they have used a principal ideal domain larger than the ring $\mathcal{O}$ of integers, which is principal only if K has class number 1. Berndt Brenken has recently told us that he has been looking at the Hecke algebras of more general almost normal inclusions from the point of view of semigroup crossed products.

Background

This paper is addressed primarily at operator algebraists, so general facts about C^*-algebras have been used freely. However, it is an attractive feature of the semigroup-crossed-product approach to Toeplitz algebras that it is relatively elementary: it requires only the basic theory of C^*-algebras and familiarity with the group C^*-algebras of discrete groups. Many of the results in the first two sections have purely algebraic analogues, involving the action α of the semigroup $\mathcal{O}^\times$ by *-endomorphisms of the group *-algebra $\mathbb{C}(K/\mathcal{O}) := \operatorname{span}\{\delta_x : x \in K/\mathcal{O}\}$.

Our notation concerning number fields is as follows. Throughout K will denote a finite extension of the rational numbers $\mathbb{Q}$, called a *number field*. Every number field has an associated *ring of integers* $\mathcal{O}$, consisting of the solutions in K of monic polynomials with coefficients in $\mathbb{Z}$; for example, $\mathbb{Z}$ is the ring of integers of $\mathbb{Q}$. We write $\mathcal{O}^\times$ for the multiplicative semigroup of nonzero integers, and $\mathcal{O}^*$ for the multiplicative group of units, or invertible elements, in $\mathcal{O}$. The only units in $\mathbb{Z}$ are ± 1, but this is certainly not true for general rings of integers: for example, real quadratic number fields have their group of units isomorphic to $\mathbb{Z}$. The field K can be recovered from $\mathcal{O}$ as its field of fractions: in other words, every number in K has the form a/b for some $a \in \mathcal{O}$ and $b \in \mathcal{O}^\times$.

The *norm* is a multiplicative homomorphism from ideals in $\mathcal{O}$ to $\mathbb{N}$, given by $N(\mathfrak{a}) = |\mathcal{O}/\mathfrak{a}|$ for an ideal $\mathfrak{a} \subseteq \mathcal{O}$. If $\mathfrak{a}$ is principally generated, so $\mathfrak{a} = a\mathcal{O}$ for some $a \in \mathcal{O}$, then this norm coincides with the absolute value of the standard number-theoretic norm $N(a)$ of the element a [11, Prop. 3.5.1]. We shall write either $N_\mathfrak{a}$ or $N(\mathfrak{a})$ to denote the norm of the ideal $\mathfrak{a}$, and for principal ideals, $N_a = |N(a)|$ will denote the norm of the ideal $a\mathcal{O}$. In §3, we shall need to use the extension of the norm to fractional ideals, but we shall discuss the key points then.

1. The semigroup dynamical system $(C^*(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$

Because $\mathcal{O}$ is a subring of K, multiplication by elements of $\mathcal{O}^\times$ gives an action of the semigroup $\mathcal{O}^\times$ as endomorphisms of the additive group $K/\mathcal{O}$. The universality of the group algebra construction allows us to lift this to an action β of $\mathcal{O}^\times$ by endomorphisms of the group C^*-algebra: thus, by definition, we have $\beta_a(\delta_x) = \delta_{ax}$ for $x \in K/\mathcal{O}$, $a \in \mathcal{O}^\times$. These are endomorphisms rather than automorphisms: as the next Lemma shows, multiplication by $a \in \mathcal{O}^\times$ is not injective at the group or group-algebra level unless a is a unit.

Lemma 1.1. *If $a \in \mathcal{O}^\times$ and $y \in K/\mathcal{O}$, the equation $ax = y$ has N_a solutions in $K/\mathcal{O}$. We write $[x : ax = y]$ for the set of solutions.*

Proof. Multiplication by a induces an isomorphism of the group $[x : ax = 0] = \frac{1}{a}\mathcal{O}/\mathcal{O}$ onto $\mathcal{O}/a\mathcal{O}$, and hence $[x : ax = 0]$ is a finite set with N_a elements. If x' is one solution of $ax' = y$, then

$$[x : ax = y] = [x : ax = ax'] = [x + x' : ax = 0] = x' + [x : ax = 0], \tag{1.1}$$

which also has N_a elements. □

When the equation $ax = y$ has more than one solution in $K/\mathcal{O}$, division by a does not give a well-defined endomorphism of $K/\mathcal{O}$. Nevertheless, one can define an endomorphism of the C^*-algebra $C^*(K/\mathcal{O})$ by averaging over the set of all solutions, and this endomorphism α_a is a right inverse for β_a. It is important to realize that the construction of α_a is not possible on $K/\mathcal{O}$ itself: one must pass to the group algebra $C^*(K/\mathcal{O})$ (or $\mathbb{C}(K/\mathcal{O})$) before the averaging makes sense.

Proposition 1.2. *Let K be a number field with ring of integers $\mathcal{O}$. The formula*

$$\alpha_a(\delta_y) = \frac{1}{N_a} \sum_{[x:ax=y]} \delta_x \tag{1.2}$$

defines an action of $\mathcal{O}^\times$ by endomorphisms of $C^*(K/\mathcal{O})$. For every $a \in \mathcal{O}^\times$, $\alpha_a(1)$ is a projection, and

$$\alpha_a(1)\alpha_b(1) = \alpha_{ab}(1) \quad \textit{whenever} \quad a\mathcal{O} + b\mathcal{O} = \mathcal{O}. \tag{1.3}$$

The action α is a right inverse for the action β defined by $\beta_a : \delta_y \mapsto \delta_{ay}$, so $\beta_a \circ \alpha_a = \mathrm{id}$, while $\alpha_a \circ \beta_a$ is multiplication by $\alpha_a(1)$.

The action α restricts to an action of $\mathcal{O}^\times$ by *-endomorphisms of the group *-algebra $\mathbb{C}(K/\mathcal{O})$.

Proof. For $y, y' \in K/\mathcal{O}$ and $a \in \mathcal{O}^\times$,

$$\begin{aligned}
\alpha_a(\delta_y)\alpha_a(\delta_{y'}) &= \left(\frac{1}{N_a}\sum_{[x:ax=y]} \delta_x\right)\left(\frac{1}{N_a}\sum_{[x':ax'=y']} \delta_{x'}\right) \\
&= \frac{1}{N_a^2}\sum_{[x:ax=y]}\sum_{[x':ax'=y']} \delta_x\delta_{x'} = \frac{1}{N_a^2}\sum_{[x:ax=y]}\sum_{[x':ax'=y']} \delta_{x+x'} \\
&= \frac{1}{N_a}\sum_{[x'':ax''=y+y']} \delta_{x''} = \alpha_a(\delta_y\delta_{y'}),
\end{aligned}$$

where the fourth equality holds because addition induces a N_a–to–one surjective map from $[x : ax = y] \times [x' : ax' = y']$ onto $[x'' : ax'' = y + y']$.

Thus $x \mapsto \alpha_a(\delta_x)$ is a homomorphism of $K/\mathcal{O}$ into $C^*(K/\mathcal{O})$, and it clearly preserves adjoints. Hence $\alpha_a(1) = \alpha_a(\delta_0)$ is a projection in the C^*-algebra $C^*(K/\mathcal{O})$, and $x \mapsto \alpha_a(\delta_x)$ is a homomorphism of $K/\mathcal{O}$ into the unitary group of the C^*-algebra $\alpha_a(1)C^*(K/\mathcal{O})\alpha_a(1)$. The universal property of $C^*(K/\mathcal{O})$ now implies that α_a extends to a homomorphism of $C^*(K/\mathcal{O})$ into itself — that is, to an endomorphism of the C^*-algebra $C^*(K/\mathcal{O})$. It follows similarly from the universal property of $\mathbb{C}(K/\mathcal{O})$ that the same formula gives *-endomorphisms α_a of $\mathbb{C}(K/\mathcal{O})$.

Next assume $a, b \in \mathcal{O}^\times$ and $z \in K/\mathcal{O}$, and calculate

$$\begin{aligned}\alpha_a(\alpha_b(\delta_z)) &= \alpha_a\left(\frac{1}{N_b}\sum_{[y:by=z]}\delta_y\right) = \frac{1}{N_a N_b}\sum_{[y:by=z]}\left(\sum_{[x:ax=y]}\delta_x\right)\\ &= \frac{1}{N_{ab}}\sum_{[x:abx=z]}\delta_x = \alpha_{ab}(\delta_z),\end{aligned}$$

where the third equality holds because $N_aN_b = N_{ab}$ and $[x : abx = z]$ is the disjoint union of the sets $[x : ax = y]$ with y ranging in $[y : by = z]$. We have now proved that α is an action by endomorphisms of $C^*(K/\mathcal{O})$, and the same calculations show that it restricts to an action on $\mathbb{C}(K/\mathcal{O})$.

To prove (1.3), multiply

$$\begin{aligned}\alpha_a(1)\alpha_b(1) &= \left(\frac{1}{N_a}\sum_{[x:ax=0]}\delta_x\right)\left(\frac{1}{N_b}\sum_{[y:by=0]}\delta_y\right)\\ &= \frac{1}{N_aN_b}\sum_{[x:ax=0]\times[y:by=0]}\delta_{x+y}\\ &= \frac{1}{N_{ab}}\sum_{[z:abz=0]}\delta_z = \alpha_{ab}(1);\end{aligned}$$

for the third equality, note that, by the Chinese Remainder Theorem, $a\mathcal{O} + b\mathcal{O} = \mathcal{O}$ implies $\mathcal{O}/ab\mathcal{O} \cong \mathcal{O}/a\mathcal{O} \times \mathcal{O}/b\mathcal{O}$, which in turn implies $\frac{1}{ab}\mathcal{O}/\mathcal{O} \cong \frac{1}{a}\mathcal{O}/\mathcal{O} \times \frac{1}{b}\mathcal{O}/\mathcal{O}$.

It is easy to check that $\beta_a(\alpha_a(\delta_y)) = \delta_y$ for any $y \in K/\mathcal{O}$. To see that $\alpha_a \circ \beta_a$ is multiplication by $\alpha_a(1)$, we compute:

$$\alpha_a(\beta_a(\delta_y)) = \frac{1}{N_a}\sum_{[x:ax=ay]}\delta_x = \frac{1}{N_a}\sum_{[x':ax'=0]}\delta_{x'+y} = \frac{1}{N_a}\left(\sum_{[x':ax'=0]}\delta_{x'}\right)\delta_y = \alpha_a(1)\delta_y,$$

where the second equality holds as in (1.1). □

Remark 1.3. Since $\beta_a \circ \alpha_a = \mathrm{id}$, α_a is injective and β_a is surjective for each $a \in \mathcal{O}^\times$. If a is a unit, $\alpha_a(1) = 1$, so $\alpha_a \circ \beta_a = \mathrm{id}$, and units act by automorphisms. Conversely, $\alpha_a(1) = 1$ only for $a \in \mathcal{O}^*$, so only units act by automorphisms. These automorphisms leave the projections $\alpha_a(1)$ fixed, because for every $a \in \mathcal{O}^\times$ and $u \in \mathcal{O}^*$, we have $\alpha_{ua}(1) = \alpha_{au}(1) = \alpha_a(\alpha_u(1)) = \alpha_a(1)$.

DEFINITION 1.4. A *covariant representation* of the system $(C^*(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$ is a pair (π, V), in which π is a unital representation of $C^*(K/\mathcal{O})$ on a Hilbert space H, and V is an isometric representation of $\mathcal{O}^\times$ on H, satisfying the covariance condition

$$\pi(\alpha_a(f)) = V_a\pi(f)V_a^* \quad \text{for } a \in \mathcal{O}^\times \text{ and } f \in C^*(K/\mathcal{O}).$$

We can use the same covariance condition to define an *algebraic covariant representation* of the system $(\mathbb{C}(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$ with values in a unital *-algebra.

This covariance condition combines with the left inverse β to give the following useful identities:

Lemma 1.5. *Suppose (π, V) is a covariant representation for $(C^*(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$. If $a, b \in \mathcal{O}^\times$ and $x \in K/\mathcal{O}$, then*

1. $V_a\pi(\delta_x) = \pi(\alpha_a(\delta_x))V_a$, $\pi(\delta_x)V_a^* = V_a^*\pi(\alpha_a(\delta_x))$,
2. $\pi(\delta_x)V_a = V_a\pi(\beta_a(\delta_x))$, $V_a^*\pi(\delta_x) = \pi(\beta_a(\delta_x))V_a^*$,
3. *and if in addition $a\mathcal{O} + b\mathcal{O} = \mathcal{O}$, then $V_a^*V_b = V_bV_a^*$.*

Proof. Since $V_a^*V_a = 1$, claim (1) is immediate from covariance. Use (1) and facts about β to compute $V_a\pi(\beta_a(\delta_x)) = V_a\pi(\delta_{ax}) = \pi(\alpha_a(\delta_{ax})V_a = \pi(\alpha_a(\beta_a(\delta_x)))V_a = \pi(\alpha_a(1)\delta_x)V_a = \pi(\delta_x)\pi(\alpha_a(1))V_a = \pi(\delta_x)V_a$, since $\alpha_a(1) = V_aV_a^*$ by covariance. The second equality in (2) is shown similarly. To see (3), multiply (1.3) by V_a^* on the left and V_b on the right. □

Example 1.6. We construct a covariant representation (λ, L) on $\ell^2(K/\mathcal{O})$, in which λ is the left regular representation of $C^*(K/\mathcal{O})$ on $\ell^2(K/\mathcal{O})$.

The isometric representation L of the semigroup $\mathcal{O}^\times$ is defined by the formula

$$L_a\epsilon_y = \frac{1}{N_a^{1/2}} \sum_{[x:ax=y]} \epsilon_x,$$

where $\{\epsilon_y : y \in K/\mathcal{O}\}$ is the usual orthonormal basis of $\ell^2(K/\mathcal{O})$. First we need to check that these are actually isometries, and for this it suffices to show that L_a maps this orthonormal basis into orthogonal unit vectors. That they are unit vectors is an easy calculation. If $ax = y \neq y' = ax'$ in $K/\mathcal{O}$ then $x \neq x'$ in $K/\mathcal{O}$, so the sums for $L_a\epsilon_y$ and $L_a\epsilon_{y'}$ are over disjoint sets, and hence orthogonal.

The same type of calculation used to show $\alpha_a \circ \alpha_b = \alpha_{ab}$ yields $L_aL_b = L_{ab}$, and one checks easily that that $L_a^*\epsilon_x = (1/N_a^{1/2})\epsilon_{ax}$, which can then be used to compute

$$\begin{aligned} L_a\lambda(\delta_x)L_a^*\epsilon_y &= \frac{1}{N_a^{1/2}} L_a\epsilon_{ay+x} = \frac{1}{N_a} \sum_{[z:az=ay+x]} \epsilon_z = \frac{1}{N_a} \sum_{[z:a(z-y)=x]} \epsilon_z \\ &= \frac{1}{N_a} \sum_{[z':az'=x]} \epsilon_{z'+y} = \lambda(\alpha_a(\delta_x))\epsilon_y. \end{aligned}$$

Therefore the pair (λ, L) is a covariant representation of the system $(C^*(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$.

Definition 1.7. Because we have just constructed a non-trivial covariant representation, we know from Proposition 2.1 of [7] that the system $(C^*(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$ has a *crossed product.* This is a C^*-algebra B generated by a *universal covariant representation* (i, v) of $(C^*(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$ in B: for every other covariant representation (π, V), there is a representation $\pi \times V$ of B such that $(\pi \times V) \circ i = \pi$ and $(\pi \times V) \circ v = V$. The triple (B, i, v) is unique up to isomorphism [7, Proposition 2.1]. Since the representation λ in the example is faithful, and $\lambda = (\lambda \times L) \circ i$, the homomorphism i is injective on $C^*(K/\mathcal{O})$.

We can similarly define the algebraic crossed product $(\mathbb{C}(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times, i, v)$ to be the *-algebra generated by a universal algebraic covariant representation. The construction of [7, Proposition 2.1] can be easily modified to show that there is such a representation.

Lemma 1.8. *The vector space* $\operatorname{span}\{v_a^* i(\delta_x) v_b : x \in K/\mathcal{O}, a, b \in \mathcal{O}^\times\}$ *is a dense* **-subalgebra of* $C^*(K/\mathcal{O}) \rtimes \mathcal{O}^\times$. *We also have*

$$\operatorname{span}\{v_a^* i(\delta_x) v_b : x \in K/\mathcal{O}, a, b \in \mathcal{O}^\times\} = \operatorname{span}\{i(\delta_x) v_a^* v_b : x \in K/\mathcal{O}, a, b \in \mathcal{O}^\times\}$$

Proof. The vector space certainly contains every $i(\delta_x)$ and v_a, and is obviously closed under taking adjoints, so it is enough to to show that the product of two spanning elements is a linear combination of such elements. To prove this let $x, y \in K/\mathcal{O}$ and $a, b, c, d \in \mathcal{O}^\times$. Then, since $v_b v_c = v_c v_b$, we have

$$\begin{aligned}
(v_a^* i(\delta_x) v_b)(v_c^* i(\delta_y) v_d) &= v_a^* i(\delta_x) v_c^* (v_b v_c)(v_b v_c)^* v_b i(\delta_y) v_d \\
&= v_a^* v_c^* i(\alpha_c(\delta_x) \alpha_{bc}(1) \alpha_b(\delta_y)) v_b v_d && \text{by Lemma 1.5 (1)} \\
&= (v_a v_c)^* i(\alpha_{bc} \circ \beta_{bc}(\alpha_c(\delta_x) \alpha_b(\delta_y))) v_b v_d && \text{by Proposition 1.2} \\
&= (v_a v_c)^* i(\alpha_{bc}(\beta_b(\delta_x)(\beta_c(\delta_y)))(v_b v_d) \\
&= (v_a v_c)^* i(\alpha_{bc}(\delta_{bx} + \delta_{cy}))(v_b v_d),
\end{aligned}$$

which we can see is in the linear span of $\{v_a^* i(\delta_x) v_b : x \in K/\mathcal{O}, a, b \in \mathcal{O}^\times\}$ by considering the formula (1.2) defining α. The last equality follows from Lemma 1.5. □

Remark 1.9. The labeling of the spanning elements by the ordered triples $(v_a, i(\delta_x), v_b)$ is not one-to-one. If $bc = ad$ and $bx = dy + n + mb/a$ for $m, n \in \mathcal{O}$, then, using Lemma 1.5(2) repeatedly,

$$\begin{aligned}
v_a^* i(\delta_x) v_b &= v_a^* v_b i(\delta_{bx}) \\
&= v_a^* v_b i(\delta_{mb/a}) i(\delta_{dy}) && \text{by assumption, since } i(\delta_n) = 1 \\
&= v_a^* i(\delta_{m/a}) v_b i(\delta_{dy}) \\
&= i(\delta_{am/a}) v_a^* v_b i(\delta_{dy}) \\
&= v_c^* v_d i(\delta_{dy}) \\
&= v_c^* i(\delta_y) v_d,
\end{aligned}$$

where the fifth equality holds because $i(\delta_m) = 1$ and $v_a^* v_b = v_a^* v_c^* v_c v_b = v_c^* v_a^* v_b v_c = v_c^* v_a^* v_a v_d = v_c^* v_d$.

From the discussion of the Hecke algebra in §2 it will follow that $v_a^* i(\delta_x) v_b = v_c^* i(\delta_y) v_d$ implies $b/a = d/c$ and $bx \equiv dy \pmod{\mathcal{O} + \frac{b}{a}\mathcal{O}}$. It will also follow that the set $\{v_a^* i(\delta_x) v_b : x \in K/\mathcal{O}, a, b \in \mathcal{O}^\times\}$ is linearly independent, hence a linear basis for the dense subalgebra $\mathbb{C}(K/\mathcal{O}) \rtimes \mathcal{O}^\times$ of $C^*(K/\mathcal{O}) \rtimes \mathcal{O}^\times$.

Proposition 1.10. *Let K be a number field with ring of integers $\mathcal{O}$. There is a strongly continuous action $\widehat{\alpha}$ of the compact group $\widehat{K^*}$ on $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ such that*

$$\widehat{\alpha}_\gamma\big(v_a^* i(\delta_x) v_b\big) = \gamma(a^{-1} b) v_a^* i(\delta_x) v_b$$

for all $\gamma \in \widehat{K^*}$, $a, b \in \mathcal{O}^\times$ and $x \in K/\mathcal{O}$; $\widehat{\alpha}$ is called the dual action.

Proof. For fixed γ, the map $w : a \mapsto \gamma(a) v_a$ gives another covariant pair (i, w), which is easily seen to be universal. Thus we can deduce from the uniqueness of the crossed product that there is an automorphism $\widehat{\alpha}_\gamma$ of $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ with the required behavior on generators. The continuity of $\gamma \mapsto \widehat{\alpha}_\gamma(c)$ is easy to check when c belongs to $\operatorname{span}\{v_a^* i(\delta_x) v_b\}$, and because automorphisms of C^*-algebras are norm-preserving, this extends to $c \in C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$. $\square$

COROLLARY 1.11. *There is a faithful positive linear map Φ of $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ onto $C^*(K/\mathcal{O})$ (strictly speaking, onto its image $i(C^*(K/\mathcal{O}))$ in the crossed product) such that*

$$\Phi\big(v_a^* i(\delta_x) v_b\big) = \begin{cases} v_a^* i(\delta_x) v_a & \text{if } b = a, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Define

$$\Phi(c) := \int_{\widehat{K^*}} \widehat{\alpha}_\gamma(c)\, d\gamma;$$

this gives a norm-decreasing projection of $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ onto the fixed-point algebra for the action $\widehat{\alpha}$, which is faithful in the sense that $\Phi(b^*b) = 0$ only if $b = 0$. Because $\int \gamma(a^{-1}b)\, d\gamma = 0$ unless $a^{-1}b = 1$, Φ has the required form on generators. The covariance of (i, v) implies that $v_a^* i(\delta_x) v_a = i(\beta_a(\delta_x)) = i(\delta_{ax})$, so Φ does indeed have range $i(C^*(K/\mathcal{O}))$. One can check by representing $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ on Hilbert space that Φ is positive (in fact, completely positive of norm 1). $\square$

Example 1.12. Composing the expectation Φ with the canonical trace $\tau : z \mapsto z(0)$ on $C^*(K/\mathcal{O})$ gives a state $\tau \circ \Phi$ on $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$. This state is faithful on positive elements because both τ and Φ are. Thus the GNS-representation $\pi_{\tau\circ\Phi}$ is a faithful representation of $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$. (We observe that when $K = \mathbb{Q}$, $\tau \circ \Phi$ is the KMS_1 state of [3, Theorem 5], which is shown there to be a factor state of type III.)

2. THE HECKE ALGEBRA OF A NUMBER FIELD

The universal property defining the crossed product $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ can be restated as a presentation in terms of generators and relations similar to the modification in [8, Corollaries 2.9 and 2.10] of [3, Proposition 18]. To do this we need to extend the definition of covariance to say that a pair (U, V) consisting of an isometric representation V of $\mathcal{O}^\times$ and a unitary representation U of $K/\mathcal{O}$ is covariant if

$$\frac{1}{N_a} \sum_{[x:ax=y]} U(x) = V_a U(y) V_a^*, \qquad \text{for } a \in \mathcal{O}^\times \text{ and } y \in K/\mathcal{O}.$$

Since $C^*(K/\mathcal{O})$ is universal for unitary representations of $K/\mathcal{O}$, a pair (U, V) is covariant in this sense precisely when (π_U, V) is a covariant representation of the dynamical system.

PROPOSITION 2.1. *The crossed product $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ is the universal C^*-algebra generated by elements $\{u(y) : y \in K/\mathcal{O}\}$, $\{v_a : a \in \mathcal{O}^\times\}$ subject to the relations:*

1. $v_a^* v_a = 1$ *for* $a \in \mathcal{O}^\times$,
2. $v_a v_b = v_{ab}$ *for* $a, b \in \mathcal{O}^\times$,
3. $u(0) = 1$, $u(x)^* = u(-x)$, $u(x)u(y) = u(x + y)$ *for* $x, y \in K/\mathcal{O}$, *and*
4. $\displaystyle\frac{1}{N_a} \sum_{[x:ax=y]} u(x) = v_a u(y) v_a^*$, *for* $a \in \mathcal{O}^\times$ *and* $x, y \in K/\mathcal{O}$.

Similarly, the algebraic crossed product $\mathbb{C}(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ *is the universal involutive algebra generated by such elements and relations.*

Proof. Relations (1) and (2) say that v is an isometric representation of $\mathcal{O}^\times$, (3) says that u is a unitary representation of $K/\mathcal{O}$, and (4) is the covariance condition. Clearly, a universal representation of the above relations is a universal covariant pair for the system $(C^*(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$, and vice versa. □

In Example 1.6 we gave a concrete representation of these relations. In this section we obtain another, by real-Ising the crossed product as a Hecke algebra, and using the regular representation of this Hecke algebra.

Recall that a subgroup Γ_0 of a group Γ is almost normal if the orbits for the left action of Γ_0 on the right coset space Γ/Γ_0 are finite. Consider the subgroup

$$\Gamma_{\mathcal{O}} = \left\{ \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} : a \in \mathcal{O} \right\} \quad \text{of}$$

$$\Gamma_K = \left\{ \begin{pmatrix} 1 & y \\ 0 & x \end{pmatrix} : x, y \in K, x \neq 0 \right\}.$$

LEMMA 2.2. $\Gamma_{\mathcal{O}}$ *is an almost normal subgroup of* Γ_K*.*

Proof. The right coset of $\gamma = \begin{pmatrix} 1 & y \\ 0 & x \end{pmatrix} \in \Gamma_K$ is $\gamma\Gamma_{\mathcal{O}} = \begin{pmatrix} 1 & y+\mathcal{O} \\ 0 & x \end{pmatrix}$, so

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \gamma\Gamma_{\mathcal{O}} = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & y+\mathcal{O} \\ 0 & x \end{pmatrix} = \begin{pmatrix} 1 & ax+y+\mathcal{O} \\ 0 & x \end{pmatrix}.$$

Thus the orbit has as many points as there are classes of $ax+y$ modulo $\mathcal{O}$. If $x = b/c$ with $b, c \in \mathcal{O}$, then $a \equiv a' \pmod{c}$ implies $ax + y \equiv a'x + y \pmod{\mathcal{O}}$, so there are at most N_c points in the orbit. □

The generalized Hecke algebra $\mathcal{H}(\Gamma_K, \Gamma_{\mathcal{O}})$ is defined in [3, §1] as a convolution *-algebra of $\Gamma_{\mathcal{O}}$-biinvariant functions on Γ_K. As a complex vector space, $\mathcal{H}(\Gamma_K, \Gamma_{\mathcal{O}})$ is the space of functions $f : \Gamma_K \to \mathbb{C}$ which are constant on double cosets, so $f(\gamma_0\gamma\gamma_0') = f(\gamma)$ for $\gamma_0, \gamma_0' \in \Gamma_{\mathcal{O}}$ and $\gamma \in \Gamma_K$, and which are supported on finitely many of these double cosets. The convolution product is

$$(f * g)(\gamma) \quad = \sum_{\gamma_1 \in \Gamma_{\mathcal{O}} \backslash \Gamma_K} f(\gamma\gamma_1^{-1}) g(\gamma_1),$$

where the sum is over left-cosets, and the involution is $f^*(\gamma) = \overline{f(\gamma^{-1})}$. With these operations, $\mathcal{H}(\Gamma_K, \Gamma_{\mathcal{O}})$ is a unital *-algebra.

It is convenient to think of $\mathcal{H}(\Gamma_K, \Gamma_{\mathcal{O}})$ as the linear span of characteristic functions of double cosets, indicated by square brackets, with the multiplication rule:

$$\begin{aligned} [\Gamma_{\mathcal{O}}\gamma_1\Gamma_{\mathcal{O}}] * [\Gamma_{\mathcal{O}}\gamma_2\Gamma_{\mathcal{O}}](\gamma) &= \sum_{\gamma' \in \Gamma_{\mathcal{O}} \backslash \Gamma_K} [\Gamma_{\mathcal{O}}\gamma_1\Gamma_{\mathcal{O}}](\gamma{\gamma'}^{-1})[\Gamma_{\mathcal{O}}\gamma_2\Gamma_{\mathcal{O}}](\gamma') \qquad (2.1) \\ &= \#\,\mathrm{LC}\left\{(\Gamma_{\mathcal{O}}\gamma_1^{-1}\Gamma_{\mathcal{O}})\gamma \cap (\Gamma_{\mathcal{O}}\gamma_2\Gamma_{\mathcal{O}})\right\}, \end{aligned}$$

where the sum is taken over representatives γ' of the left cosets $\Gamma_{\mathcal{O}} \backslash \Gamma_K$, and $\#\,\mathrm{LC}$ counts the number of left cosets in a left-invariant subset of Γ_K. The last equality holds because the term of the sum corresponding to a left coset γ' is 0 unless

$\gamma\gamma'^{-1} \in \Gamma_{\mathcal{O}}\gamma_1\Gamma_{\mathcal{O}}$ and $\gamma' \in \Gamma_{\mathcal{O}}\gamma_2\Gamma_{\mathcal{O}}$, in which case it is 1. Involution is determined by conjugate-linearity and $[\Gamma_{\mathcal{O}}\gamma\Gamma_{\mathcal{O}}]^* = [\Gamma_{\mathcal{O}}\gamma^{-1}\Gamma_{\mathcal{O}}]$, and the unit is $[\Gamma_{\mathcal{O}}]$.

Consider the maps $\mu : \mathcal{O}^\times \to \mathcal{H}(\Gamma_K, \Gamma_{\mathcal{O}})$ and $e : K \to \mathcal{H}(\Gamma_K, \Gamma_{\mathcal{O}})$ defined by

$$\mu_a = \frac{1}{N_a^{1/2}}\left[\Gamma_{\mathcal{O}}\begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}\Gamma_{\mathcal{O}}\right] \tag{2.2}$$

$$e(r) = \left[\Gamma_{\mathcal{O}}\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}\Gamma_{\mathcal{O}}\right]. \tag{2.3}$$

The map e factors through $K/\mathcal{O}$ because $\Gamma_{\mathcal{O}}\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}\Gamma_{\mathcal{O}} = \begin{pmatrix} 1 & r+\mathcal{O} \\ 0 & 1 \end{pmatrix}$, and the same notation will be used for the corresponding map of $K/\mathcal{O}$ into $\mathcal{H}(\Gamma_K, \Gamma_{\mathcal{O}})$. The following generalization of [3, Proposition 18] shows that the Hecke algebra is generated by these elements, and that they are universal generators. More precisely, it says that the pair (e, μ) is covariant and that $\pi_e \times \mu$ is a *-algebra isomorphism of $\mathbb{C}(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ onto $\mathcal{H}(\Gamma_K, \Gamma_{\mathcal{O}})$.

Theorem 2.3. *Let K be a number field with ring of integers $\mathcal{O}$. The elements μ_a and $e(x)$ defined in* (2.2) *and* (2.3)*, with $a \in \mathcal{O}^\times$ and $x \in K/\mathcal{O}$, generate the Hecke algebra $\mathcal{H}(\Gamma_K, \Gamma_{\mathcal{O}})$, and satisfy the relations*

- $\mathcal{H}1.$ $\mu_a^*\mu_a = 1$ *for* $a \in \mathcal{O}^\times$,
- $\mathcal{H}2.$ $\mu_a\mu_b = \mu_{ab}$ *for* $a, b \in \mathcal{O}^\times$,
- $\mathcal{H}3.$ $e(0) = 1,\ e(x)^* = e(-x)$ *and* $e(x)e(y) = e(x+y)$ *for* $x, y \in K/\mathcal{O}$, *and*
- $\mathcal{H}4.$ $\frac{1}{N_a}\sum_{[x:ax=y]} e(x) = \mu_a e(y)\mu_a^*$, *for* $a \in \mathcal{O}^\times$ *and* $y \in K/\mathcal{O}$.

Moreover, $\mathcal{H}(\Gamma_K, \Gamma_{\mathcal{O}})$ is the universal *-algebra over $\mathbb{C}$ with these generators and relations; it is spanned by the set $\{\mu_a^* e(x)\mu_b : a, b \in \mathcal{O}^\times, x \in K\}$.

Proof. To prove $(\mathcal{H}3)$, first observe that

$$\Gamma_{\mathcal{O}}\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}\Gamma_{\mathcal{O}} = \Gamma_{\mathcal{O}}\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}\Gamma_{\mathcal{O}} = \begin{pmatrix} 1 & r+\mathcal{O} \\ 0 & 1 \end{pmatrix},$$

so for these elements, left cosets, right cosets and double cosets coincide. Let $r, s \in K$, $\gamma = \begin{pmatrix} 1 & y \\ 0 & x \end{pmatrix} \in \Gamma_K$, and compute as in (2.1):

$$\begin{aligned} e(r)e(s)(\gamma) &= \left[\begin{pmatrix} 1 & r+\mathcal{O} \\ 0 & 1 \end{pmatrix}\right] * \left[\begin{pmatrix} 1 & s+\mathcal{O} \\ 0 & 1 \end{pmatrix}\right](\gamma) \\ &= \#\,\mathrm{LC}\left\{\begin{pmatrix} 1 & -r+\mathcal{O} \\ 0 & 1 \end{pmatrix}\gamma \cap \begin{pmatrix} 1 & s+\mathcal{O} \\ 0 & 1 \end{pmatrix}\right\} \\ &= \#\,\mathrm{LC}\left\{\begin{pmatrix} 1 & y-rx+x\mathcal{O} \\ 0 & x \end{pmatrix} \cap \begin{pmatrix} 1 & s+\mathcal{O} \\ 0 & 1 \end{pmatrix}\right\} \\ &= \begin{cases} 1 & \text{if } x = 1 \text{ and } y \equiv r+s \pmod{\mathcal{O}} \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

because if $x = 1$ and $y - r \equiv s \pmod{\mathcal{O}}$, the intersection is the (single) left coset $\begin{pmatrix} 1 & s+\mathcal{O} \\ 0 & 1 \end{pmatrix}$. Thus $e(r)e(s) = e(r+s)$. The remaining identities are easily verified.

To see ($\mathcal{H}1$) and ($\mathcal{H}2$), notice that $\Gamma_{\mathcal{O}}\begin{pmatrix}1 & 0\\ 0 & a\end{pmatrix}\Gamma_{\mathcal{O}} = \begin{pmatrix}1 & \mathcal{O}\\ 0 & a\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & a\end{pmatrix}\Gamma_{\mathcal{O}}$, so the support of μ_a is a right coset, and the support of μ_a^* is the left coset $\Gamma_{\mathcal{O}}\begin{pmatrix}1 & \frac{1}{a}\mathcal{O}\\ 0 & \frac{1}{a}\end{pmatrix}$. Thus, for $\gamma = \begin{pmatrix}1 & y\\ 0 & x\end{pmatrix}$, we have

$$\begin{aligned}
\mu_a^*\mu_a(\gamma) &= \frac{1}{N_a}\left[\begin{pmatrix}1 & \frac{1}{a}\mathcal{O}\\ 0 & \frac{1}{a}\end{pmatrix}\right] * \left[\begin{pmatrix}1 & \mathcal{O}\\ 0 & a\end{pmatrix}\right](\gamma)\\
&= \frac{1}{N_a}\#\,\mathrm{LC}\left\{\begin{pmatrix}1 & \mathcal{O}\\ 0 & a\end{pmatrix}\gamma \cap \begin{pmatrix}1 & \mathcal{O}\\ 0 & a\end{pmatrix}\right\}\\
&= \frac{1}{N_a}\#\,\mathrm{LC}\left\{\begin{pmatrix}1 & y+x\mathcal{O}\\ 0 & ax\end{pmatrix} \cap \begin{pmatrix}1 & \mathcal{O}\\ 0 & a\end{pmatrix}\right\}\\
&= \begin{cases}1 & \text{if } x = 1 \text{ and } y \in \mathcal{O}\\ 0 & \text{otherwise,}\end{cases}
\end{aligned}$$

because if $x = 1$ and $y \in \mathcal{O}$ the intersection $\begin{pmatrix}1 & \mathcal{O}\\ 0 & a\end{pmatrix}$ contains exactly N_a left cosets. This proves $\mu_a^*\mu_a = \left[\begin{pmatrix}1 & \mathcal{O}\\ 0 & 1\end{pmatrix}\right] = [\Gamma_{\mathcal{O}}] = 1$. A similar computation proves ($\mathcal{H}2$).

Before proving the covariance condition ($\mathcal{H}4$), we compute $\mu_a e(r)$:

$$\begin{aligned}
\mu_a e(r)(\gamma) &= \frac{1}{N_a^{1/2}}\#\,\mathrm{LC}\left\{\begin{pmatrix}1 & \frac{1}{a}\mathcal{O}\\ 0 & \frac{1}{a}\end{pmatrix}\gamma \cap \begin{pmatrix}1 & r+\mathcal{O}\\ 0 & 1\end{pmatrix}\right\}\\
&= \frac{1}{N_a^{1/2}}\#\,\mathrm{LC}\left\{\begin{pmatrix}1 & y+\frac{x}{a}\mathcal{O}\\ 0 & \frac{1}{a}x\end{pmatrix} \cap \begin{pmatrix}1 & r+\mathcal{O}\\ 0 & 1\end{pmatrix}\right\}\\
&= \begin{cases}1/N_a^{1/2} & \text{if } x = a \text{ and } y \equiv r \pmod{\mathcal{O}}\\ 0 & \text{otherwise.}\end{cases}
\end{aligned}$$

Thus $\mu_a e(r) = \frac{1}{N_a^{1/2}}\left[\begin{pmatrix}1 & r+\mathcal{O}\\ 0 & a\end{pmatrix}\right]$ and

$$\begin{aligned}
\mu_a e(r)\mu_a^*(\gamma) &= \frac{1}{N_a}\left[\begin{pmatrix}1 & r+\mathcal{O}\\ 0 & a\end{pmatrix}\right] * \left[\begin{pmatrix}1 & \frac{1}{a}\mathcal{O}\\ 0 & \frac{1}{a}\end{pmatrix}\right](\gamma)\\
&= \frac{1}{N_a}\#\,\mathrm{LC}\left\{\begin{pmatrix}1 & -\frac{r}{a}+\frac{1}{a}\mathcal{O}\\ 0 & \frac{1}{a}\end{pmatrix}\gamma \cap \begin{pmatrix}1 & \frac{1}{a}\mathcal{O}\\ 0 & \frac{1}{a}\end{pmatrix}\right\}\\
&= \frac{1}{N_a}\#\,\mathrm{LC}\left\{\begin{pmatrix}1 & y-\frac{rx}{a}+\frac{x}{a}\mathcal{O}\\ 0 & \frac{x}{a}\end{pmatrix} \cap \begin{pmatrix}1 & \frac{1}{a}\mathcal{O}\\ 0 & \frac{1}{a}\end{pmatrix}\right\}\\
&= \begin{cases}1/N_a & \text{if } x = 1 \text{ and } y - r/a \in \frac{1}{a}\mathcal{O}\\ 0 & \text{otherwise.}\end{cases}
\end{aligned}$$

This gives

$$\mu_a e(r)\mu_a^* = \frac{1}{N_a}\left[\begin{pmatrix}1 & \frac{1}{a}(r+\mathcal{O})\\ 0 & 1\end{pmatrix}\right],$$

which implies ($\mathcal{H}4$) because the right-hand-side is the sum of N_a characteristic functions of double cosets, one for each class in $r/a + (1/a)\mathcal{O} \pmod{\mathcal{O}}$; in other words,

$$\mu_a e(r)\mu_a^* = \frac{1}{N_a}\sum_{[x:ax=r]}\left[\begin{pmatrix} 1 & x+\mathcal{O} \\ 0 & 1\end{pmatrix}\right] = \frac{1}{N_a}\sum_{[x:ax=r]} e(x) = \pi_e(\alpha_a(\delta_r)).$$

Now that we have verified ($\mathcal{H}1$)–($\mathcal{H}4$), the universal property of the algebraic crossed product gives a *-algebra homomorphism $\pi_e \times \mu$ of $\mathbb{C}(K/\mathcal{O}) \rtimes \mathcal{O}^\times$ into the Hecke algebra $\mathcal{H}(\Gamma_K, \Gamma_\mathcal{O})$, and it only remains to prove that $\pi_e \times \mu$ is one-to-one and onto.

Consider a single monomial $\mu_a^* e(r)\mu_b$. A computation similar to the one above gives

$$\mu_a^* e(r)(\gamma) = \frac{1}{N_a^{1/2}}\left[\begin{pmatrix} 1 & r+\frac{1}{a}\mathcal{O} \\ 0 & \frac{1}{a}\end{pmatrix}\right],$$

and further calculation shows

$$\mu_a^* e(r)\mu_b(\gamma) = \frac{1}{N_{ab}^{1/2}}\#\,\mathrm{LC}\left\{\begin{pmatrix} 1 & y-rb+\frac{b}{a}\mathcal{O} \\ 0 & ax\end{pmatrix}\cap\begin{pmatrix} 1 & \mathcal{O} \\ 0 & b\end{pmatrix}\right\}.$$

Thus we must have $x = b/a$ and $y \in rb + \frac{b}{a}\mathcal{O} + \mathcal{O}$. Since $\begin{pmatrix} 1 & \mathcal{O} \\ 0 & b\end{pmatrix}$ is *not* a (single) left coset, we must count carefully to find the number of left cosets in this intersection. We notice, first, that $ab\mathcal{O} \subseteq b\mathcal{O}\cap a\mathcal{O} \subseteq a\mathcal{O}$, $\frac{b}{a}\mathcal{O}\cap\mathcal{O}$ is an ideal in $\mathcal{O}$ and $(\frac{b}{a}\mathcal{O}\cap\mathcal{O})/b\mathcal{O} \cong (b\mathcal{O}\cap a\mathcal{O})/ab\mathcal{O}$, and, second, that $a\mathcal{O}/(b\mathcal{O}\cap a\mathcal{O}) \cong \mathcal{O}/(\frac{b}{a}\mathcal{O}\cap\mathcal{O})$, so that $|a\mathcal{O}/(b\mathcal{O}\cap a\mathcal{O})| = N(\frac{b}{a}\mathcal{O}\cap\mathcal{O})$. From the isomorphism theorems we have

$$|a\mathcal{O}/(b\mathcal{O}\cap a\mathcal{O})|\,|(b\mathcal{O}\cap a\mathcal{O})/ab\mathcal{O}| = |a\mathcal{O}/ab\mathcal{O}| = |\mathcal{O}/b\mathcal{O}| = |N(b)| = N_b,$$

and from the multiplicativity of the norm, we deduce that the number of left cosets is $N_b/N(\frac{b}{a}\mathcal{O}\cap\mathcal{O})$. We divide by $N_{ab}^{1/2}$ and manipulate to get

$$\mu_a^* e(r)\mu_b = \frac{N(\frac{b}{a})^{1/2}}{N(\frac{b}{a}\mathcal{O}\cap\mathcal{O})}\left[\begin{pmatrix} 1 & rb+\frac{b}{a}\mathcal{O}+\mathcal{O} \\ 0 & \frac{b}{a}\end{pmatrix}\right].$$

The support of the right hand side is a single double-coset. To see this, multiply one of its elements on the left and on the right by $\Gamma_\mathcal{O}$ to get

$$\begin{pmatrix} 1 & \mathcal{O} \\ 0 & 1\end{pmatrix}\begin{pmatrix} 1 & rb \\ 0 & \frac{b}{a}\end{pmatrix}\begin{pmatrix} 1 & \mathcal{O} \\ 0 & 1\end{pmatrix} = \begin{pmatrix} 1 & rb+\frac{b}{a}\mathcal{O}+\mathcal{O} \\ 0 & \frac{b}{a}\end{pmatrix}.$$

Since every double coset has this form, and since $N(\frac{b}{a})^{1/2} \neq 0$, the linear span of the elements $\mu_a^* e(r)\mu_b$ is all of $\mathcal{H}(\Gamma_K, \Gamma_\mathcal{O})$. Moreover, if two such elements $\mu_a^* e(x)\mu_b$ and $\mu_c^* e(y)\mu_d$ do not have disjoint support, they are supported on the same double coset, in which case $b/a = d/c$ and $\mu_a^* e(x)\mu_b = \mu_c^* e(y)\mu_d$. Thus the set $\{\mu_a^* e(x)\mu_b : a, b \in \mathcal{O}^\times\ x \in K/\mathcal{O}\}$ is linearly independent, because distinct elements have disjoint support.

Since the representation $\pi_e \times \mu$ maps $\{v_a^* u(x) v_b : x \in K/\mathcal{O}, a, b \in \mathcal{O}^\times\}$ injectively onto a linear basis for the Hecke algebra, it follows that $\{v_a^* u(x) v_b : x \in K/\mathcal{O}$ and $a, b \in \mathcal{O}^\times\}$ is a linear basis for the algebraic crossed product and that

$$\pi_e \times \mu : \mathbb{C}(K/\mathcal{O}) \rtimes \mathcal{O}^\times \to \mathcal{H}(\Gamma_K, \Gamma_\mathcal{O})$$

is a *-algebra isomorphism. The result now follows from Proposition 2.1. □

The Hecke algebra $\mathcal{H}(\Gamma_K, \Gamma_\mathcal{O})$ acts as convolution operators on the Hilbert space $\ell^2(\Gamma_\mathcal{O}\backslash\Gamma_K)$, and then the Hecke C^*-algebra $C^*(\Gamma_K, \Gamma_\mathcal{O})$ is by definition the closure of $\mathcal{H}(\Gamma_K, \Gamma_\mathcal{O})$ in the operator norm, [3, Proposition 3], [2]. Thus, the generators $e(r)$ and μ_a, viewed as unitaries and isometries on $\ell^2(\Gamma_\mathcal{O}\backslash\Gamma_K)$, give a covariant representation (π_e, μ) of $(C^*(K/\mathcal{O}), u, v)$ such that $C^*(\Gamma_K, \Gamma_\mathcal{O}) = (\pi_e \times \mu)(C^*(K/\mathcal{O}) \rtimes \mathcal{O}^\times)$. It will follow from our main theorem in §4 that this *Hecke representation* is faithful; i.e. that the Hecke C^*-algebra is the universal C^*-algebra of the relations $(\mathcal{H}1)$–$(\mathcal{H}4)$.

We can also establish directly that the Hecke representation is faithful by embedding the faithful representation of Example 1.12 as a subrepresentation. Indeed, the subspace of $\ell^2(\Gamma_\mathcal{O}\backslash\Gamma_K)$ consisting of biinvariant functions is invariant under the Hecke representation (π_e, μ), and the corresponding subrepresentation turns out to be the GNS-representation of the state $\tau \circ \Phi$.

Proposition 2.4. *The representation of the Hecke algebra as convolution operators on $\ell^2(\Gamma_\mathcal{O}\backslash\Gamma_K/\Gamma_\mathcal{O})$ is unitarily equivalent to the GNS-representation of $\tau \circ \Phi$.*

Proof. By uniqueness of the GNS-representation, it is enough to show that the vector $[\Gamma_\mathcal{O}] \in \ell^2(\Gamma_\mathcal{O}\backslash\Gamma_K/\Gamma_\mathcal{O})$ is cyclic for the left convolution action of $\mathcal{H}(\Gamma_K, \Gamma_\mathcal{O})$ and that the corresponding vector state $\omega_{\Gamma_\mathcal{O}}$ is equal to $\omega \circ \Phi$. Since $[\Gamma_\mathcal{O}]$ is an identity for convolution, its cyclic component contains every biinvariant function supported on finitely many double cosets; this proves that $[\Gamma_\mathcal{O}]$ is cyclic.

To show that $\omega_{\Gamma_\mathcal{O}} = \tau \circ \Phi$, notice first that, because the fixed point algebra of the dual action $\hat\alpha$ of $\widehat{K^\times}$ is exactly $C^*(K/\mathcal{O})$, any state ω of $C^*(K/\mathcal{O})$ has a *unique* $\hat\alpha$-invariant extension to $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$, namely $\omega \circ \Phi$. So it suffices to prove that the vector state $\omega_{\Gamma_\mathcal{O}}$ is $\hat\alpha$-invariant and agrees with τ on $C^*(K/\mathcal{O})$. If $a \neq b$, then the support of $\mu_a^* e(r)\mu_b[\Gamma_\mathcal{O}]$ is disjoint from $\Gamma_\mathcal{O}$, and hence $\omega_{\Gamma_\mathcal{O}}(\mu_a^* e(r)\mu_b) = \langle \mu_a^* e(r)\mu_b[\Gamma_\mathcal{O}], [\Gamma_\mathcal{O}], \rangle = 0$. Similarly, if $r \neq 0$ the support of $e(r)[\Gamma_\mathcal{O}]$ is disjoint from $[\Gamma_\mathcal{O}]$, and hence $\omega_{\Gamma_\mathcal{O}}(e(r)) = \langle e(r)[\Gamma_\mathcal{O}], [\Gamma_\mathcal{O}]\rangle = 0$. Since we trivially have $\omega_{\Gamma_\mathcal{O}}(e(0)) = 1$, this proves that $\omega_{\Gamma_\mathcal{O}}$ is $\hat\alpha$-invariant and agrees with τ on $C^*(K/\mathcal{O})$, as required. ☐

Corollary 2.5. *Let K be a number field with ring of integers $\mathcal{O}$. Then the Hecke representation $\pi_e \times \mu$ is faithful on $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ and the Hecke C^*-algebra $C^*(\Gamma_K, \Gamma_\mathcal{O})$ is the universal C^*-algebra of the relations $(\mathcal{H}1)$–$(\mathcal{H}4)$.*

3. Characters of $K/\mathcal{O}$

In [8] the character $\varkappa(r) = \exp(2\pi i r)$ gave an embedding of $\mathbb{Q}/\mathbb{Z}$ in $\mathbb{T}$ which was essential to the characterization of faithful covariant representations. There is no such embedding in general:

Lemma 3.1. *If K is a nontrivial extension of $\mathbb{Q}$, there are no injective characters of $K/\mathcal{O}$.*

Proof. Suppose that K is an extension of degree $[K : \mathbb{Q}] = n > 1$, and choose an integer $a \in \mathbb{Z} \cap \mathcal{O}^\times$ with $a \neq \pm 1$. Then the subgroup $\frac{1}{a}\mathcal{O}/\mathcal{O}$ of $K/\mathcal{O}$ has order $N_a = a^n$ [11, 2.6(3)]. On the other hand, every $x \in \mathcal{O}$ satisfies $x = ax/a = 0$ in $\frac{1}{a}\mathcal{O}/\mathcal{O}$, so the order of $\chi(x/a)$ divides a for every character χ. Thus $\chi(\frac{1}{a}\mathcal{O}/\mathcal{O})$ is a subgroup of the a^{th}-roots of unity and χ cannot be injective. ☐

For $\chi \in (K/\mathcal{O})\widehat{}$ and $b \in \mathcal{O}$, define a character χ^b on $K/\mathcal{O}$ by $\chi^b(x) := \chi(bx)$. Our key technical Lemma says that for every number field K there exists $\chi \in (K/\mathcal{O})\widehat{}$ such that $\{\chi^b : b \in \mathcal{O}\}$ is dense in $(K/\mathcal{O})\widehat{}$ (Corollary 3.5, Lemma 3.6); these characters play the role of the injective characters of $\mathbb{Q}/\mathbb{Z}$. We begin by recording a general fact.

Lemma 3.2. *Let χ be a character on $K/\mathcal{O}$, and let $a, b \in \mathcal{O}^\times$. Then*

$$\sum_{[x:ax=0]} \chi(bx) = 0 \text{ if and only if } \chi(bx) \neq 1 \text{ for some } x \in [x : ax = 0]. \tag{3.1}$$

Proof. The set $\{\chi(bx) : ax = 0\}$ is a group of roots of unity, and hence, unless this group is trivial, its elements sum to zero. □

In dealing with semigroup crossed products $A \rtimes_\alpha S$, one often needs to know that $\prod_{a\in F}(1 - \alpha_a(1))$ is nonzero for every finite set of elements F of S (see [7, Theorem 3.7], for example). In the present setting, something stronger is needed. The problem is that $\alpha_a(1)\alpha_b(1)$ is not necessarily of the form $\alpha_c(1)$ for $c \in \mathcal{O}^\times$. To get around this, we would like to make sense of $\alpha_\mathfrak{a}(1)$ for ideals $\mathfrak{a}$ in $\mathcal{O}$, in such a way that $\alpha_a(1)\alpha_b(1) = \alpha_\mathfrak{a}(1)$ with $\mathfrak{a}$ the not-necessarily-principal ideal generated by a and b. The ideals in $\mathcal{O}$ form a semigroup including $\mathcal{O}^\times/\mathcal{O}^*$ as the subsemigroup of principal ideals, but we have been unable to find a suitable action α of this semigroup on $C^*(K/\mathcal{O})$. However, we can define projections $P_\mathfrak{a}$ which have the properties we require of $\alpha_\mathfrak{a}(1)$. Once we have established these properties in Proposition 3.4, we can show the existence of the required characters on $K/\mathcal{O}$ (Corollary 3.5, Lemma 3.6).

We need some basic facts about *fractional ideals.* A fractional ideal $\mathfrak{f}$ of a number field K is a nonzero finitely-generated $\mathcal{O}$-submodule of K such that $d\mathfrak{f} \subset \mathcal{O}$ for some $d \in \mathcal{O}^\times$. Ideals in $\mathcal{O}$ are certainly fractional ideals, with $d = 1$; these are called *integral ideals* when it is necessary to distinguish them. Products and inverses of fractional ideals are defined by

$$\mathfrak{f}\mathfrak{g} = \{\sum_{i=1}^n f_i g_i : f_i \in \mathfrak{f}, g_i \in \mathfrak{g}\}$$
$$\mathfrak{f}^{-1} = \{x \in K : x\mathfrak{f} \subset \mathcal{O}\},$$

and are fractional ideals too. Since the ring of integers $\mathcal{O}$ is a Dedekind domain, these operations make the set of fractional ideals into a multiplicative group $\mathcal{I}_K$ with identity element the ideal $\mathcal{O}$; moreover, every element in $\mathcal{I}_K$ can be factored uniquely into a product of integer powers of prime ideals in $\mathcal{O}$. Hence $\mathcal{I}_K$ is a free Abelian group with the set $\mathcal{P}$ of prime ideals as generators [11, Theorem 3.4.3].

The intersection $\mathfrak{f} \cap \mathfrak{g}$ of two fractional ideals, which is sometimes denoted $[\mathfrak{f}, \mathfrak{g}]$, is a greatest lower bound in terms of ideal inclusion; similarly, $\mathfrak{f} + \mathfrak{g}$, which is sometimes denoted $(\mathfrak{f}, \mathfrak{g})$, is the least upper bound. The notation of lcm and gcd is meaningful; if $\mathfrak{f}$ and $\mathfrak{g}$ are two fractional ideals with factorizations

$$\mathfrak{f} = \prod_{\mathfrak{p}\in\mathcal{P}} \mathfrak{p}^{n_\mathfrak{p}(\mathfrak{f})} \quad \text{and} \quad \mathfrak{g} = \prod_{\mathfrak{p}\in\mathcal{P}} \mathfrak{p}^{n_\mathfrak{p}(\mathfrak{g})},$$

then

$$[\mathfrak{f}, \mathfrak{g}] = \mathfrak{f} \cap \mathfrak{g} = \prod_{\mathfrak{p}\in\mathcal{P}} \mathfrak{p}^{\max(n_\mathfrak{p}(\mathfrak{f}), n_\mathfrak{p}(\mathfrak{g}))},$$

and

$$(\mathfrak{f}, \mathfrak{g}) = \mathfrak{f} + \mathfrak{g} = \prod_{\mathfrak{p}\in\mathcal{P}} \mathfrak{p}^{\min(n_\mathfrak{p}(\mathfrak{f}), n_\mathfrak{p}(\mathfrak{g}))}.$$

Notice that with these factorizations, if $\mathfrak{f}$ is integral, all the exponents $n_{\mathfrak{p}}$ are nonnegative, and if $\mathfrak{f}$ is the inverse of an integral ideal, $n_{\mathfrak{p}} \leq 0$ for all $\mathfrak{p}$. Thus any fractional ideal can be written as $\mathfrak{f} = \frac{\mathfrak{a}}{\mathfrak{b}}$, with $\mathfrak{a}, \mathfrak{b} \subseteq \mathcal{O}$, and we can define the norm of a fractional ideal by $N(\mathfrak{f}) = N(\mathfrak{a})/N(\mathfrak{b})$ [6, pp. 17,24]. However, if $\mathfrak{f}$ is not integral this norm no longer represents a cardinality.

If $\mathfrak{a}$ is an integral ideal, then $\mathfrak{a}^{-1}$ contains $\mathcal{O}$. Let $d \in \mathcal{O}$ be an integer such that $d\mathfrak{a}^{-1} \subseteq \mathcal{O}$. Since we trivially have $d\mathcal{O} \subseteq d\mathfrak{a}^{-1}$, the isomorphism theorems give

$$|\mathcal{O}/d\mathcal{O}| = \left|\mathcal{O}/d\mathfrak{a}^{-1}\right| \left|d\mathfrak{a}^{-1}/d\mathcal{O}\right|;$$

since $d\mathfrak{a}^{-1}/d\mathcal{O} \cong \mathfrak{a}^{-1}/\mathcal{O}$, we deduce that

$$\left|\mathfrak{a}^{-1}/\mathcal{O}\right| = \left|d\mathfrak{a}^{-1}/d\mathcal{O}\right| = \frac{N_d}{N(d\mathfrak{a}^{-1})} = N(\mathfrak{a}).$$

Lemma 3.3. *Suppose* $\mathfrak{a}$ *and* $\mathfrak{b}$ *are integral ideals in* $\mathcal{O}$. *Then*

$$0 \to (\mathfrak{a}+\mathfrak{b})^{-1}/\mathcal{O} \xrightarrow[x \mapsto (x,-x)]{} \mathfrak{a}^{-1}/\mathcal{O} \times \mathfrak{b}^{-1}/\mathcal{O} \xrightarrow[(x,y)\mapsto x+y]{} (\mathfrak{a} \cap \mathfrak{b})^{-1}/\mathcal{O} \to 0$$

is an exact sequence of finite Abelian groups.

Proof. From the factorization into prime ideals it is easy to see that $\mathfrak{a}^{-1} + \mathfrak{b}^{-1} = (\mathfrak{a} \cap \mathfrak{b})^{-1}$ and $\mathfrak{a}^{-1} \cap \mathfrak{b}^{-1} = (\mathfrak{a}+\mathfrak{b})^{-1}$. Hence addition gives a natural surjective homomorphism $(x,y) \in \mathfrak{a}^{-1} \times \mathfrak{b}^{-1} \mapsto x+y \in (\mathfrak{a} \cap \mathfrak{b})^{-1}$ with kernel $\{(x,-x) : x \in (\mathfrak{a}+\mathfrak{b})^{-1}\}$. Taking quotients by $\mathcal{O}$ gives the sequence. □

We are now ready to define the projections $P_{\mathfrak{a}}$ in $C^*(K/\mathcal{O})$.

Proposition 3.4. *For each integral ideal* $\mathfrak{a}$ *in* $\mathcal{O}$ *let*

$$P_{\mathfrak{a}} = \frac{1}{N(\mathfrak{a})} \sum_{x \in \mathfrak{a}^{-1}/\mathcal{O}} \delta_x, \tag{3.2}$$

where the sum is taken over any set of representatives of $\mathfrak{a}^{-1}/\mathcal{O}$. *Then*

(i) $P_{(a)} = \alpha_a(1)$ *for every* $a \in \mathcal{O}^{\times}$,
(ii) $P_{\mathfrak{a}}$ *is a projection for every* $\mathfrak{a}$,
(iii) $P_{\mathfrak{a}} \geq P_{\mathfrak{b}}$ *whenever* $\mathfrak{a}|\mathfrak{b}$ *(i.e. whenever* $\mathfrak{b} \subset \mathfrak{a}$*)*,

and, for every finite collection $\{\mathfrak{a}_i\}_{1\leq i\leq n}$ *of integral ideals,*

(iv) $\prod_i P_{\mathfrak{a}_i} = P_{\cap_i \mathfrak{a}_i}$, *and*
(v) $\prod_i (1 - P_{\mathfrak{a}_i}) \neq 0$ *whenever* $\mathfrak{a}_i \neq \mathcal{O}$ *for* $1 \leq i \leq n$.

Proof. Claim (i) is verified directly from the definition. Since multiplication and intersection are associative operations, to prove (iv) it is enough to consider two ideals $\mathfrak{a}$ and $\mathfrak{b}$:

$$\begin{aligned} P_{\mathfrak{a}}P_{\mathfrak{b}} &= \frac{1}{N(\mathfrak{a})N(\mathfrak{b})} \sum_{x\in\mathfrak{a}^{-1}/\mathcal{O}} \sum_{y\in\mathfrak{b}^{-1}/\mathcal{O}} \delta_{x+y} \\ &= \frac{N(\mathfrak{a}+\mathfrak{b})}{N(\mathfrak{a})N(\mathfrak{b})} \sum_{z\in(\mathfrak{a}\cap\mathfrak{b})^{-1}/\mathcal{O}} \delta_z \\ &= \frac{1}{N(\mathfrak{a}\cap\mathfrak{b})} \sum_{z\in(\mathfrak{a}\cap\mathfrak{b})^{-1}/\mathcal{O}} \delta_z \\ &= P_{\mathfrak{a}\cap\mathfrak{b}}, \end{aligned}$$

where the second equality holds by Lemma 3.3. Since $\mathfrak{a}^{-1}/\mathcal{O}$ contains $-x$ whenever it contains x, $P_\mathfrak{a}$ is self adjoint, and setting $\mathfrak{a} = \mathfrak{b}$ in (iv) gives $P_\mathfrak{a}^2 = P_\mathfrak{a}$, proving (ii). If $\mathfrak{b}|\mathfrak{a}$ then $\mathfrak{a} \cap \mathfrak{b} = \mathfrak{b}$, so (iii) follows from (iv).

It remains to prove (v). Observe first that replacing each $\mathfrak{a}_i$ by one of its prime factors gives a smaller projection because of (iii); repeated primes are irrelevant because the $P_{\mathfrak{a}_i}$ are idempotents. Thus it suffices to prove that $\prod_{\mathfrak{a}\in F}(1 - P_\mathfrak{a}) \neq 0$ for any finite set F of distinct prime ideals. Multiplying out and using (iv) gives

$$\prod_{\mathfrak{a}\in F}(1 - P_\mathfrak{a}) = \sum_{A\subset F}\prod_{\mathfrak{a}\in A}(-P_\mathfrak{a}) = \sum_{A\subset F}(-1)^{|A|}P_{\cap A},$$

where $\cap A$ indicates the intersection of all the members of A, which in this case equals their product because they are all prime. This projection is in $\mathbb{C}(K/\mathcal{O})$, and, viewing it as a function on $K/\mathcal{O}$, it makes sense to evaluate it at $0 \in K/\mathcal{O}$:

$$\begin{aligned}
\prod_{\mathfrak{a}\in F}(1 - P_\mathfrak{a})(0) &= \sum_{A\subset F}(-1)^{|A|}P_{\cap A}(0) \\
&= \sum_{A\subset F}(-1)^{|A|}\frac{1}{N(\cap A)}\sum_{x\in(\cap A)^{-1}/\mathcal{O}}\delta_x(0) \\
&= \sum_{A\subset F}\prod_{\mathfrak{a}\in A}(-\frac{1}{N(\mathfrak{a})}), \qquad \text{because } N(\cap A) = \prod_{\mathfrak{a}\in A}N(\mathfrak{a}), \\
&= \prod_{\mathfrak{a}\in F}(1 - \frac{1}{N(\mathfrak{a})}) \neq 0,
\end{aligned}$$

because $N(\mathfrak{a}) > 1$ for every integral ideal $\mathfrak{a} \neq \mathcal{O}$. □

Corollary 3.5. *Let $f \mapsto \widehat{f}$ denote the Fourier transform isomorphism of $C^*(K/\mathcal{O})$ onto $C(\widehat{K/\mathcal{O}})$. Then*

$$\mathcal{X}_K := \bigcap\{\operatorname{supp}\widehat{1 - P_\mathfrak{a}} : \mathfrak{a} \textit{ is a nontrivial ideal in } \mathcal{O}\}$$

is a nonempty compact G_δ subset of $\widehat{K/\mathcal{O}}$.

Proof. The space $\widehat{K/\mathcal{O}}$ is compact, and the family $\{\operatorname{supp}(1 - P_\mathfrak{a})\widehat{\ }\}$ has the finite intersection property by Proposition 3.4(v). □

The following lemma shows that the characters in $\mathcal{X}_K$ have the required properties.

Lemma 3.6. *Let $\chi \in \widehat{K/\mathcal{O}}$. Then*

1. *$\chi \in \mathcal{X}_K$ if and only if $\chi(\mathfrak{a}^{-1}/\mathcal{O}) \neq \{1\}$ for every non-trivial ideal $\mathfrak{a} \subseteq \mathcal{O}$,*
2. *if $\chi \in \mathcal{X}_K, a, b \in \mathcal{O}^\times$, and $\chi(bx) = 1$ for all $x \in \frac{1}{a}\mathcal{O}/\mathcal{O}$, then $a|b$, and*
3. *if $\chi \in \mathcal{X}_K$, then $\{\chi^b : b \in \mathcal{O}\}$ is dense in $\widehat{K/\mathcal{O}}$.*

Proof. Suppose $\chi \in \mathcal{X}_K$. By the definition of the set $\mathcal{X}_K$, $\hat{P}_\mathfrak{a}(\chi) \neq 1$, so it must be zero, which means $\sum_{x\in\mathfrak{a}^{-1}/\mathcal{O}}\chi(x) = 0$. Equivalently, the group $\chi(\mathfrak{a}^{-1}/\mathcal{O})$ of roots of unity is non-trivial by (3.1), giving (1). To see (2), note that

$$\tfrac{1}{a}(a\mathcal{O} + b\mathcal{O}) = \tfrac{1}{a}\{ax + by : x, y \in \mathcal{O}\} = \{x + \tfrac{b}{a}y : x, y \in \mathcal{O}\}.$$

Suppose a does not divide b, and set $\mathfrak{a}^{-1} = \frac{1}{a}(a\mathcal{O} + b\mathcal{O})$: this makes sense since by dividing ideals we can compute

$$\tfrac{1}{a}\left(a\mathcal{O} + b\mathcal{O}\right) = \tfrac{1}{a}\left(\tfrac{ab\mathcal{O}}{a\mathcal{O}\cap b\mathcal{O}}\right) = (a)^{-1}\left(\tfrac{a\mathcal{O}\cap b\mathcal{O}}{ab\mathcal{O}}\right)^{-1},$$

and so $\mathfrak{a} = (a\mathcal{O}\cap b\mathcal{O})/b\mathcal{O}$ is an integral ideal. If $\chi \in \mathcal{X}_K$, then from (1) we have

$$\chi(\{\frac{by}{a} : y \in \mathcal{O}\}) = \chi(\{x + \frac{by}{a} : x, y \in \mathcal{O}\}) = \chi(\mathfrak{a}^{-1}) \neq \{1\},$$

so (2) is proved.

Let $\chi \in \mathcal{X}_K$. The map $b \mapsto \chi^b$ from $\mathcal{O}$ to the characters on $K/\mathcal{O}$ is a group homomorphism. We claim that the homomorphism $b \mapsto \chi^b|_{\frac{1}{a}\mathcal{O}/\mathcal{O}}$ has kernel $a\mathcal{O}$. We see that a is in the kernel, since $\chi^a(\frac{1}{a}\mathcal{O}) = \chi(\mathcal{O}) = \{1\}$. Suppose b is in the kernel. Then $\chi(bx) = 1$ for all $x \in \frac{1}{a}\mathcal{O}/\mathcal{O}$, so (2) implies that $a|b$; thus $b \in a\mathcal{O}$, and the claim is true. Thus we have an injective homomorphism of $\mathcal{O}/a\mathcal{O}$ into $(\frac{1}{a}\mathcal{O}/\mathcal{O})\widehat{\ }$, and since these are finite Abelian groups of the same cardinality N_a, the homomorphism must also be surjective. Thus every character on $\frac{1}{a}\mathcal{O}/\mathcal{O}$ is the restriction of some χ^b. Since $K/\mathcal{O} = \cup\{\frac{1}{a}\mathcal{O}/\mathcal{O} : a \in \mathcal{O}^\times\}$, we have

$$\widehat{K/\mathcal{O}} = \varprojlim \widehat{\tfrac{1}{a}\mathcal{O}/\mathcal{O}},$$

and we can deduce that $\{\chi^b : b \in \mathcal{O}\}$ is dense in $\widehat{K/\mathcal{O}}$. □

Remark 3.7. The referee suggested that it should also be possible to prove the existence of characters with the required properties using Fourier analysis on the adele group A of K, as in [6]. In fact, this method is used by Harari and Leichtnam [5]. The approach presented here is more elementary, and in particular bypasses the application of the strong approximation theorem.

The characters in $\mathcal{X}_K$ will play a very important rôle in the proof of our main theorem. We can also use them to construct new covariant representations of the system $(C^*(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$ involving the usual Toeplitz representation T of $\mathcal{O}^\times$ on $\ell^2(\mathcal{O}^\times)$, which is defined in terms of the usual basis $\{\varepsilon_b : b \in \mathcal{O}^\times\}$ for $\ell^2(\mathcal{O}^\times)$ by $T_a(\varepsilon_b) := \varepsilon_{ab}$.

Proposition 3.8. *Suppose $\chi \in \mathcal{X}_K$. Then $\tau_\chi(x) : \varepsilon_b \mapsto \chi^b(x)\varepsilon_b$ extends to a faithful representation of $C^*(K/\mathcal{O})$ such that the pair (τ_χ, T) is covariant.*

Proof. The operator $\tau_\chi(\delta_x)$ is multiplication by the circle-valued function $b \mapsto \chi^b(x)$ on $\ell^2(\mathcal{O}^\times)$, so τ_χ is a unitary representation of $K/\mathcal{O}$; we use the same symbol for the corresponding representation of $C^*(K/\mathcal{O})$. For $f \in C^*(K/\mathcal{O})$, $\tau_\chi(f)$ is multiplication by the function $b \mapsto \widehat{f}(\chi^b)$, and since $\{\chi^b : b \in \mathcal{O}^\times\}$ is dense in $(K/\mathcal{O})\widehat{\ }$ by Lemma 3.6, τ_χ is faithful.

To check the covariance condition, fix $b \in \mathcal{O}^\times$. Compute first

$$T_a\tau_\chi(y)T_a^*\varepsilon_b = \begin{cases} T_a\tau_\chi(y)\varepsilon_{b/a} & \text{if } a|b \\ 0 & \text{if } a \not| b \end{cases} = \begin{cases} \chi((b/a)y)\varepsilon_b & \text{if } a|b \\ 0 & \text{if } a \not| b, \end{cases}$$

and then

$$\tau_\chi(\alpha_a(y))\varepsilon_b = \frac{1}{N_a}\sum_{[x:ax=y]} \tau_\chi(x)\varepsilon_b = \left(\frac{1}{N_a}\sum_{[x:ax=y]} \chi(bx)\right)\varepsilon_b.$$

Let z be a fixed element of $[z : az = y]$. Then

$$\begin{aligned}\frac{1}{N_a}\sum_{[x:ax=y]}\chi(bx) &= \frac{1}{N_a}\sum_{[x':ax'=0]}\chi(b(x'+z)) = \chi(bz)\frac{1}{N_a}\sum_{[x':ax'=0]}\chi(bx')\\ &= \begin{cases}\chi(bz) & \text{if } a|b\\ 0 & \text{if } a\not| b,\end{cases}\end{aligned}$$

by Lemma 3.6(2) and (3.1). Since $a|b$ implies $\chi(bz) = \chi((b/a)az) = \chi((b/a)y)$, covariance follows. □

4. Representations of the crossed product

In this section we prove our main theorem — the characterization of faithful representations of the crossed product — and then discuss the various specific representations we have constructed earlier.

Theorem 4.1. *Let K be a number field with ring of integers $\mathcal{O}$. A covariant representation $\pi\times V$ of $C^*(K/\mathcal{O})\rtimes_\alpha\mathcal{O}^\times$ is faithful if and only if π is faithful.*

The strategy of the proof is familiar: the crux is to show that deleting the terms with $a\neq b$ from finite sums $\sum_{a,b\in F}\pi(f)V_a^*V_b$ gives a norm-decreasing expectation of $\pi\times V(C^*(K/\mathcal{O})\rtimes_\alpha\mathcal{O}^\times)$ onto $\pi(C^*(K/\mathcal{O}))$. For this, we want a projection $Q=\pi(q)$ such that compressing by Q kills the off-diagonal terms while retaining the norm of the remaining sum of diagonal terms (see Lemma 4.3 below). The presence of invertible elements (units) in the semigroup $\mathcal{O}^\times$ makes this trickier than it was in [8], and we begin with a lemma which will help deal with units. Both the next two lemmas depend crucially on the characters constructed in the previous section.

Lemma 4.2. *Suppose $\chi\in\mathcal{X}_K$, $c\in\mathcal{O}^\times$ and H is a finite set of units in $\mathcal{O}$. Then there is a projection $q\in C^*(K/\mathcal{O})$ such that $q\alpha_u(q)=0$ for all $u\in H$ and $\widehat{q}(\chi^c)=1$.*

Proof. We begin by observing that the units in $\mathcal{O}$ act as automorphisms of $C^*(K/\mathcal{O})$ (the inverse of α_u is $\beta_{u^{-1}}$), and hence α induces an action of $\mathcal{O}^*$ on the spectrum $(K/\mathcal{O})\widehat{}$ of $C^*(K/\mathcal{O})$. Indeed, we have $u\cdot\theta(x) := \theta(\alpha_u^{-1}(x)) = \theta(ux) = \theta^u(x)$ for every θ in $(K/\mathcal{O})\widehat{}$. We claim that $\mathcal{O}^*$ acts freely on the set $\{\chi^b : b\in\mathcal{O}^\times\}$. To see why, suppose $u\in\mathcal{O}^*$ satisfies $u\cdot\chi^b=\chi^b$ — or, equivalently, $\chi^{ub}=\chi^b$. Then for all $x\in K/\mathcal{O}$, we have

$$1=\chi^{ub}(x)\chi^b(x)^{-1}=\chi((u-1)bx).$$

By Lemma 3.6, this implies that every $a\in\mathcal{O}^\times$ divides $(u-1)b$, and this is only possible if $u=1$. This justifies the claim.

The claim implies that the characters $\{u\cdot\chi^c=\chi^{uc} : u\in H\}$ are distinct elements of $(K/\mathcal{O})\widehat{}$. Since the discrete group $K/\mathcal{O}=\cup_a\frac{1}{a}\mathcal{O}/\mathcal{O}$ is a directed union of finite subgroups, the dual $(K/\mathcal{O})\widehat{}$ is a topological inverse limit of finite groups, and hence is a totally disconnected compact Hausdorff space. Thus we can find a compact neighborhood N of χ^c such that $(u\cdot N)\cap N=\emptyset$ for all $u\in H$. Its characteristic function $1_N\in C((K/\mathcal{O})\widehat{})$ is the Fourier transform of a projection $q\in C^*(K/\mathcal{O})$ with the required properties. □

Recall from Lemma 1.8 that the crossed product $C^*(K/\mathcal{O})\rtimes_\alpha\mathcal{O}^\times$ is the closed linear span of $\{i(f)v_a^*v_b : f\in\mathbb{C}(K/\mathcal{O}) \text{ and } a,b\in\mathcal{O}^\times\}$.

LEMMA 4.3. *Let* $\sum_{a,b\in F} i(f_{a,b})v_a^* v_b$ *be a finite linear combination with* $f_{a,b} \in C^*(K/\mathcal{O})$*, and let* $\epsilon > 0$*. Then there exists a projection* $q = q(\epsilon) \in C^*(K/\mathcal{O})$ *such that*

$$i(q)i(f_{a,b})v_a^* v_b i(q) = 0 \qquad \text{if } a \neq b, \text{ and} \tag{4.1}$$

$$\Big\| q\Big(\sum f_{a,a}\Big)q\Big\| \geq \Big\|\sum f_{a,a}\Big\| - \epsilon. \tag{4.2}$$

Proof. Let $\chi \in \mathcal{X}_K$ and let $g = \sum f_{a,a} \in C^*(K/\mathcal{O})$. By Lemma 3.6(3) there exists $c \in \mathcal{O}^\times$ such that $|\widehat{g}(\chi^c)| \geq \|\widehat{g}\| - \epsilon$. Consider the projection

$$q_1 = \alpha_c(1)\prod_{a\not| b}(1 - \beta_b \circ \alpha_{ac}(1))\prod_{b\not| a}(1 - \beta_a \circ \alpha_{bc}(1)).$$

If $a \in F$ is not associate to $b \in F$ then either $a \not| b$ or $b \not| a$. Suppose first $b \not| a$. Then $i(q_1)i(f_{a,b})v_a^* v_b i(q_1)$ has a factor

$$\begin{aligned}
& i((\alpha_c(1) - \alpha_c(1)\beta_a(\alpha_{bc}(1))))v_a^* v_b i(\alpha_c(1)) = \\
&= v_a^* i((\alpha_{ac}(1) - \alpha_{ac}(1)\alpha_a \circ \beta_a(\alpha_{bc}(1)))\alpha_{bc}(1))v_b \quad \text{by Lemma 1.5(1),} \\
&= v_a^* i((\alpha_{ac}(1) - \alpha_{ac}(1)\alpha_a(1)\alpha_{bc}(1))\alpha_{bc}(1))v_b \\
&= v_a^* i((\alpha_{ac}(1) - \alpha_{ac}(1)\alpha_{bc}(1))\alpha_{bc}(1))v_b \\
&= 0.
\end{aligned}$$

The case $a \not| b$ reduces to this one by taking adjoints.

We now consider $H := \{u \in \mathcal{O}^* \setminus \{1\} : \text{there exists } a \in F \text{ with } ua \in F\}$. By Lemma 4.2, there is a projection q_2 such that $q_2\alpha_u(q_2) = 0$ for all $u \in H$ and $\widehat{q_2}(\chi^c) = 1$. We claim that the projection $q := q_1 q_2$ has the required properties. Indeed, the calculation in the previous paragraph shows that $i(q)v_a^* v_b i(q) = 0$ when $a, b \in F$ are not associate. If a is associate to b, then $b = ua$ for some $u \in H$, and $v_a^* v_b = v_u$; now the property $q_2\alpha_u(q_2) = 0$ forces $i(q)v_a^* v_b i(q) = i(q)v_u i(q) = 0$.

By construction, χ^c is in the support of $\widehat{q_2}$, so to finish the proof of (4.2) we need to show that $\widehat{q_1}(\chi^c) = 1$. Since χ^c is always in the support of $\alpha_c(1)\widehat{\ }$, it suffices to prove that $(\beta_a \circ \alpha_{bc}(1))\widehat{\ }(\chi^c) = 0$ whenever $b \not| a$ in $\mathcal{O}^\times$.

$$\begin{aligned}
(\beta_a \circ \alpha_{bc}(1))\widehat{\ }(\chi^c) &= \frac{1}{N_{bc}} \sum_{[x:bcx=0]} \widehat{\beta_a(\delta_x)}(\chi^c) \\
&= \frac{1}{N_{bc}} \sum_{[x:bcx=0]} \chi(cax).
\end{aligned}$$

By Lemma 3.6(2), at least one of the summands is $\neq 1$, because bc does not divide ac. Thus the sum vanishes by (3.1). $\square$

Recall from Corollary 1.11 that we have a faithful linear map $\Phi : C^*(K/\mathcal{O}) \rtimes \mathcal{O}^\times \to C^*(K/\mathcal{O})$, constructed by averaging over the compact orbits of $(K^*)\widehat{\ }$.

PROPOSITION 4.4. *Let* (π, V) *be covariant for* $(C^*(K/\mathcal{O}), \mathcal{O}^\times, \alpha)$*. If* π *is faithful, the map*

$$\phi : \pi(f)V_a^* V_b \mapsto \begin{cases} \pi(f) & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$$

extends by linearity and continuity to a projection of norm 1 *from* $C^*(\pi, V)$ *onto* $C^*(\pi)$*, such that the following diagram commutes*

$$\begin{array}{ccc} C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times & \xrightarrow{\pi\times V} & C^*(\pi, V) \\ \Big\downarrow \Phi & & \Big\downarrow \phi \\ C^*(K/\mathcal{O}) & \xrightarrow{\pi} & C^*(\pi). \end{array} \tag{4.3}$$

Proof. Let $\sum_{a,b\in F} \pi(f_{a,b})V_a^* V_b$ be a linear combination of the spanning monomials and fix $\epsilon > 0$. Let q be the projection from Lemma 4.3, and take $Q := \pi(q)$. Since π is faithful, it is isometric. Thus

$$\begin{aligned} \Big\| \sum_{a,b\in F} \pi(f_{a,b})V_a^* V_b \Big\| &\geq \Big\| Q \sum_{a,b\in F} \pi(f_{a,b})V_a^* V_b Q \Big\| \\ &= \Big\| \sum_a Q\pi(f_{a,a})V_a^* V_a Q \Big\| \\ &= \Big\| \sum_a q f_{a,a} q \Big\| \\ &\geq \Big\| \sum_a f_{a,a} \Big\| - \epsilon \\ &= \Big\| \sum_a \pi(f_{a,a}) \Big\| - \epsilon. \end{aligned}$$

Since ϵ is arbitrary, this gives the existence of the contractive projection ϕ. That the diagram commutes is easily verified on the spanning set. □

Proof. [Proof of Theorem 4.1.] Since there is a covariant representation (λ, L) with λ faithful, and this representation factors through (i, v), i must be faithful. Thus if $\pi \times V$ is faithful, so is $\pi = (\pi \times V) \circ i$. For the other direction, suppose π is faithful and $\pi \times V(b) = 0$. Then $\pi(\Phi(b^*b)) = \phi(\pi \times V)(b^*b) = 0$, and the faithfulness of Φ on positive elements implies $b = 0$. □

Next we consider the various covariant representations of $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$:

1. The representation $\lambda \times L$ on $\ell^2(K/\mathcal{O})$ (Example 1.6).
2. The GNS-representation associated to the state $\tau \circ \Phi$ on $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$, which is already known to be faithful (Example 1.12).
3. The Hecke representation on $\ell^2(\Gamma_\mathcal{O} \backslash \Gamma_K)$ (see §2).
4. The representations $\tau_\chi \times T$ from Proposition 3.8.
5. A one-dimensional representation: the trivial character on $K/\mathcal{O}$ and the trivial representation of $\mathcal{O}^\times$ on $\mathbb{C}$ form a covariant pair.

Corollary 4.5. *The representations* (1), (3) *and* (4) *of* $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ *are all faithful.*

As things stand, it is not obvious that these representations are different. In fact (λ, L) is quite different: the dual action is not unitarily implemented. Our proof of this shows more: the representations $\{\lambda \times \gamma L : \gamma \in (K^*)\widehat{\ }\}$ are a family of mutually inequivalent irreducible representations.

Proposition 4.6. *Suppose that* U *is a non-zero bounded operator on* $\ell^2(K/\mathcal{O})$*, and that there exists* $\gamma \in \widehat{K^*}$ *such that*

1. $U\lambda_x = \lambda_x U$ *for all* $x \in K/\mathcal{O}$, *and*
2. $UL_a = \gamma(a)L_a U$ *for all* $a \in \mathcal{O}^\times$.

Then U is a scalar multiple of 1 *and* $\gamma = 1$.

Proof. Let $u_x := (U\varepsilon_0|\varepsilon_x)$. Then $\sum_{x\in K/\mathcal{O}} |u_x|^2 = \|U\varepsilon_0\|^2 < \infty$. Condition (1) implies

$$\begin{aligned} (U\varepsilon_y|\varepsilon_x) &= (U\lambda_y\varepsilon_0|\varepsilon_x) = (\lambda_y U\varepsilon_0|\varepsilon_x) = (U\varepsilon_0|\lambda_y^*\varepsilon_x) && (4.4)\\ &= (U\varepsilon_0|\lambda_{-y}\varepsilon_x) = (U\varepsilon_0|\varepsilon_{x-y}) = u_{x-y}. && (4.5)\end{aligned}$$

(We think of $U \sim \sum u_y\lambda_y$ as the Fourier series of U, which by (1) belongs to the maximal Abelian algebra $\lambda(K/\mathcal{O})''$.) We claim that, for each fixed $n \in \mathbb{N} \subset \mathcal{O}$ and each $x \in K/\mathcal{O}$, we have

$$\sum_{[y:ny=x]} u_y = \gamma(n)u_x.$$

To see this, we use (2) and calculate:

$$\begin{aligned} \gamma(n)u_x &= \big(\gamma(n)U\varepsilon_0|\varepsilon_x\big) = \big(L_n^*UL_n\varepsilon_0|\varepsilon_x\big) = \big(UL_n\varepsilon_0|L_n\varepsilon_x\big)\\ &= \Big(U\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_{i/n}\Big)\Big|\frac{1}{\sqrt{n}}\sum_{[y:ny=x]}\varepsilon_x\Big)\\ &= \frac{1}{n}\sum_i\sum_{[y:ny=x]} u_{y-i/n}.\end{aligned}$$

Now $\{y - i/n : ny = x, 1 \le i \le n\}$ is n copies of $[y : ny = x]$, so

$$\gamma(n)u_x = \frac{1}{n}\sum_{[y:ny=x]} nu_y = \sum_{[y:ny=x]} u_y,$$

as claimed.

Now suppose that $u_x \neq 0$ for some $x \neq 0$, and fix $n \in \mathbb{N}$. Recall that the ℓ^2- and ℓ^1-norms on $\mathbb{C}^n$ are related by $\|z\|_2 \ge \|z\|_1/\sqrt{n}$. Thus the claim implies that

$$|u_x| = \Big|\sum_{[y:ny=x]} u_y\Big| \le \sum_{[y:ny=x]} |u_y| \le \sqrt{n}\Big(\sum_{[y:ny=x]} |u_y|^2\Big)^{1/2}.$$

We deduce that

$$\sum_{y\in K/\mathcal{O}} |u_y|^2 \ge \sum_{n\in\mathbb{N}}\Big(\sum_{[y:ny=x]} |u_y|^2\Big) \ge \sum_n \frac{|u_x|^2}{n} = |u_x|^2\Big(\sum_n \frac{1}{n}\Big) = \infty,$$

contradicting $\sum |u_y|^2 = \|U\varepsilon_0\|^2 < \infty$. □

Corollary 4.7. *The representations $\{(\lambda, \chi L) : \chi \in \widehat{K^*}\}$ are irreducible and mutually inequivalent.*

Proof. For the first assertion, take $\gamma = 1$ in the proposition, and multiply both sides by $\chi(a)$. To see that $(\lambda, \chi_1 L)$ is not equivalent to $(\lambda, \chi_2 L)$, apply the proposition with $\gamma = \chi_1^{-1}\chi_2$. □

Corollary 4.8. *The automorphisms in the dual action $\widehat{\alpha}$ of $\widehat{K^*}$ on $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ are not implemented by unitaries in the representation $\lambda \times L$.*

Remark 4.9. That the dual action is not implemented distinguishes the representations $\lambda \times \gamma L$ from the others in the list. For example, because the state $\omega \circ \Phi$ is invariant under the dual action $\widehat{\alpha}$, there is a unitary representation U of $(K^*)\widehat{}$ on $H_{\omega\circ\Phi}$ such that $(\pi_{\omega\circ\Phi}, U)$ is a covariant representation of $(C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times, (K^*)\widehat{}, \widehat{\alpha})$. It is also easy to check that the representation $U : (K^*)\widehat{} \to B(\ell^2(\mathcal{O}^\times))$ defined by $U_\gamma \varepsilon_a = \gamma(a)\varepsilon_a$ gives a covariant representation $(\tau \times T, U)$.

To see that the dual action is unitarily implemented in the Hecke representation, define $U : (K^*)\widehat{} \to B(\ell^2(\Gamma_{\mathcal{O}} \backslash \Gamma_K))$ by

$$U_\gamma : \left[\begin{pmatrix} 1 & y + x\mathcal{O} \\ 0 & x \end{pmatrix}\right] \mapsto \gamma(x) \left[\begin{pmatrix} 1 & y + x\mathcal{O} \\ 0 & x \end{pmatrix}\right].$$

The necessary relations $U_\gamma e(r) = e(r)U_\gamma$ and $U_\gamma \mu_a = \gamma(a)\mu_a U_\gamma$ follow easily by observing that

$$\operatorname{supp}\left(e(r) * \left[\begin{pmatrix} 1 & y + x\mathcal{O} \\ 0 & x \end{pmatrix}\right]\right) \subset \begin{pmatrix} 1 & * \\ 0 & x \end{pmatrix}, \text{ and}$$

$$\operatorname{supp}\left(\mu_a * \left[\begin{pmatrix} 1 & y + x\mathcal{O} \\ 0 & x \end{pmatrix}\right]\right) \subset \begin{pmatrix} 1 & * \\ 0 & ax \end{pmatrix}.$$

Remark 4.10. The representation $\lambda \times L$ is the GNS-representation corresponding to the vector state $\phi : c \mapsto (\lambda \times L(c)\varepsilon_0|\varepsilon_0)$. Since $\tau \circ \Phi = \int_{\widehat{K^*}} \phi \circ \widehat{\alpha}_\gamma \, d\gamma$, it is tempting to guess that $\pi_{\tau\circ\Phi}$ is the direct integral of the representations $\lambda \times \gamma L = (\lambda \times L) \circ \widehat{\alpha}_\gamma$. However, because each $\lambda \times \gamma L$ is irreducible, the direct integral representation on $L^2((K^*)\widehat{}, \ell^2(K/\mathcal{O}))$ has commutant $L^\infty((K^*)\widehat{})$, and is therefore type I. On the other hand, in the case $K = \mathbb{Q}$, $\tau \circ \Phi$ is the KMS$_1$-state described in [3, §1], and this is known to be a factor state of type III$_1$ [3, Theorem 5].

5. Fields of class number 1

The ideal class group of a field K is the quotient of the group F of fractional ideals by the subgroup P of principally generated ideals; it is a finite Abelian group whose cardinality is called the class number h_K of the field [11, §4.3]. The group of principal ideals is always isomorphic to $K^*/\mathcal{O}^*$, so we have an exact sequence

$$1 \to \mathcal{O}^* \to K^* \to F \to F/P \to 1$$

of Abelian groups. Since fractional ideals factor uniquely as products of prime ideals, when $h_K = |F/P| = 1$, $K^*/\mathcal{O}^*$ is the free Abelian group generated by the prime ideals. It is possible in this case to choose a multiplicative section S in $\mathcal{O}^\times$ consisting of one associate for each class in $\mathcal{O}^\times$: select an arbitrary prime generator from each prime ideal, and take S to consist of 1 and the products of the selected generators.

Throughout this section, K will be a number field with $h_K = 1$, and S will be such a subsemigroup of $\mathcal{O}^\times$. The semigroup S is lattice ordered in the sense of [10, 7], with $a \vee b$ defined to be the unique representative in S of the ideal generated generated by a and b. Restricting α to S gives another semigroup dynamical system $(C^*(K/\mathcal{O}), S, \alpha)$ associated to a number field of class number 1.

In the case of $K = \mathbb{Q}$, selecting the positive primes gives the section $\mathbb{N}^*$, and the dynamical system $(C^*(\mathbb{Q}/\mathbb{Z}), \mathbb{N}^*, \alpha)$ is the one studied in [8]. In fact S is always non-canonically isomorphic to $\mathbb{N}^* \cong \oplus_{p\in\mathcal{P}}\mathbb{N}$, so in some sense the dynamical systems

$(C^*(K/\mathcal{O}), S, \alpha)$ involve different actions of the same lattice-ordered semigroup. However, the inclusion of $\mathbb{Z}$ in $\mathcal{O}$ induces a canonical inclusion of $\mathbb{N}^*$ in S, which takes each prime generator of $\mathbb{N}^*$ to the unique product in S of (the representatives in S of) its prime factors, and this is not an isomorphism unless $K = \mathbb{Q}$.

The pairs (λ, L) and (τ_χ, T) restrict to covariant representations of $(C^*(K/\mathcal{O}), S, \alpha)$ which are faithful on $C^*(K/\mathcal{O})$, so it follows from [7, Proposition 2.1] that the system has a unique crossed product $C^*(K/\mathcal{O}) \rtimes_\alpha S$. The following version of our main theorem is a direct generalization of [8, Theorem 3.7].

THEOREM 5.1. *Suppose K is a number field with $h_K = 1$, and $(C^*(K/\mathcal{O}), S, \alpha)$ is the dynamical system constructed above. Then a representation $\pi \times V$ is faithful on $C^*(K/\mathcal{O}) \rtimes_\alpha S$ if and only if π is faithful.*

This theorem can be proved by modifying the proof of Theorem 4.1. The crossed product $C^*(K/\mathcal{O}) \rtimes_\alpha S$ carries a dual action of $(K^*)\widehat{\ }$, and averaging over this dual action gives a faithful expectation of $C^*(K/\mathcal{O}) \rtimes_\alpha S$ onto $C^*(K/\mathcal{O})$ (as in Proposition 1.10 and Corollary 1.11). The analogue of Lemma 4.3 is easier: if $\sum_{a,b\in F} f_{a,b} v_a^* v_b$ is a finite sum in $C^*(K/\mathcal{O}) \rtimes_\alpha S$, then no two different elements of F are associates, and we can take for q the projection q_1 constructed in the first paragraph of the proof of Lemma 4.3. Now the proofs of Proposition 4.4 and Theorem 4.1 carry over verbatim, giving Theorem 5.1.

It is interesting to note that Theorem 5.1 is substantially deeper than in the special case $K = \mathbb{Q}$ [8, Theorem 3.7]; it depends crucially on the existence of characters χ such that $\{\chi^b : b \in \mathcal{O}\}$ is dense in $(K/\mathcal{O})\widehat{\ }$, which was much easier in the case of $\mathbb{Q}$ (compare Corollary 3.5 and Lemma 3.6(3) with [8, Lemma 2.5]).

Remark 5.2. The crossed product $C^*(K/\mathcal{O}) \rtimes_\alpha S$ is the Hecke C^*-algebra $C^*(\Gamma_S, \Gamma_\mathcal{O})$ of the almost normal inclusion

$$\Gamma_\mathcal{O} = \begin{pmatrix} 1 & \mathcal{O} \\ 0 & 1 \end{pmatrix} \subset \Gamma_S = \begin{pmatrix} 1 & K \\ 0 & SS^{-1} \end{pmatrix}.$$

To see this, note that $\Gamma_\mathcal{O}\backslash\Gamma_S/\Gamma_\mathcal{O}$ is a subset of $\Gamma_\mathcal{O}\backslash\Gamma_K/\Gamma_\mathcal{O}$, so $\mathcal{H}(\Gamma_S, \Gamma_\mathcal{O})$ naturally embeds in $\mathcal{H}(\Gamma_K, \Gamma_\mathcal{O})$. As in the proof of Theorem 2.3, the characteristic function of every double coset is $\mu_a^* e(x) \mu_b$ for some $a, b \in S$ and $x \in K/\mathcal{O}$, so $\mathcal{H}(\Gamma_S, \Gamma_\mathcal{O})$ is generated by $\{\mu_a : a \in S\}$ and $\{e(x) : x \in K/\mathcal{O}\}$; they still satisfy the relations $(\mathcal{H}1)$–$(\mathcal{H}4)$ for $a, b \in S$, and are linearly independent because they have disjoint support. Hence $\mathcal{H}(\Gamma_S, \Gamma_\mathcal{O})$ is the universal *-algebra with such generators and relations. Theorem 5.1 therefore implies that the completion $C^*(\Gamma_S, \Gamma_\mathcal{O})$ is isomorphic to $C^*(K/\mathcal{O}) \rtimes_\alpha S$.

Remark 5.3. Because the semigroup S is lattice-ordered, we can write down an alternative spanning set for the crossed product $C^*(K/\mathcal{O}) \rtimes_\alpha S$:

$$C^*(K/\mathcal{O}) \rtimes_\alpha S = \overline{\operatorname{span}}\{i(x) v_a v_b^* : x \in K/\mathcal{O},\ a, b \in S \text{ with } (a, b) = 1\}.$$

To see this, first note that because ideals are principal, Proposition 3.4 yields

$$\alpha_a(1)\alpha_b(1) = \alpha_{a\vee b}(1),$$

which is equivalent to $v_a v_a^* v_b v_b^* = v_{a\vee b} v_{a\vee b}^*$. Multiplying on the left by v_a^*, right by v_b gives

$$v_a^* v_b = v_a^* v_{a\vee b} v_{a\vee b}^* v_b = v_{a^{-1}(a\vee b)} v_{b^{-1}(a\vee b)}^*;$$

this suffices to prove the claim because $(a^{-1}(a \vee b), b^{-1}(a \vee b)) = 1$.

Remark 5.4. It follows from Theorem 5.1 that $C^*(K/\mathcal{O}) \rtimes_\alpha S$ embeds as a subalgebra of $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$. In fact we can recover $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ from this subalgebra by taking the crossed product by the action γ of $\mathcal{O}^*$ satisfying

$$\gamma_u(i(f)v_a^* v_b) = i(\alpha_u(f))v_a^* v_b.$$

To see this, first observe that the unitary elements v_u implement the automorphisms γ_u, so there is a homomorphism π of $(C^*(K/\mathcal{O}) \rtimes_\alpha S) \rtimes \mathcal{O}^*$ into $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$. On the other hand, because $\mathcal{O}^\times$ is the direct product of $\mathcal{O}^*$ and S, we can combine the embeddings of $\mathcal{O}^*$ and S in $(C^*(K/\mathcal{O}) \rtimes_\alpha S) \rtimes \mathcal{O}^*$ into one homomorphism of $\mathcal{O}^\times$, which is covariant with the embedding of $C^*(K/\mathcal{O})$, and hence gives a homomorphism ρ of $C^*(K/\mathcal{O}) \rtimes_\alpha \mathcal{O}^\times$ into the iterated crossed product. It is easy to check that π and ρ are inverses of each other.

References

[1] S. Adji, M. Laca, M. Nilsen and I. Raeburn, *Crossed products by semigroups of endomorphisms and the Toeplitz algebras of ordered groups*, Proc. Amer. Math. Soc. 122 (1994), 1133–1141.

[2] M.W. Binder, *Induced factor representations of discrete groups and their types*, J. Functional Analysis 115 (1993), 294–312.

[3] J.-B. Bost and A. Connes, *Hecke algebras, Type III factors and phase transitions with spontaneous symmetry breaking in number theory*, Selecta Math. (New Series) 1 (1995), 411–457.

[4] J. Cuntz, *Simple C^*-algebras generated by isometries*, Comm. Math. Phys. 57 (1977), 173–185.

[5] D. Harari and E. Leichtnam, Extension du phénomène de brisure spontanée de symétrie de Bost-Connes au cas de corps globaux quelconques, preprint, June 1996.

[6] S. Lang, *Algebraic Number Theory*, Springer-Verlag, New York, 1986.

[7] M. Laca and I. Raeburn, *Semigroup crossed products and the Toeplitz algebras of non-Abelian groups*, J. Functional Analysis 139 (1996), 415–440.

[8] M. Laca and I. Raeburn, *A semigroup crossed product arising in number theory*, J. London Math. Soc., to appear.

[9] G. J. Murphy, *Crossed products of C^*-algebras by endomorphisms*, Integral Equations & Operator Theory 24 (1996), 298–319.

[10] A. Nica, *C^*-algebras generated by isometries and Wiener-Hopf operators*, J. Operator Theory 27 (1992), 17–52.

[11] P. Samuel, *Algebraic Theory of Numbers*, Hermann, Paris, 1970.

Jane Arledge
Department of Mathematics
University of Newcastle
NSW 2308
Australia
arledge@math.newcastle.edu.au

Marcelo Laca
Department of Mathematics
University of Newcastle
NSW 2308
Australia
marcelo@math.newcastle.edu.au

Iain Raeburn
Department of Mathematics
University of Newcastle
NSW 2308
Australia
iain@math.newcastle.edu.au

Bivariante K-Theorie für lokalkonvexe Algebren und der Chern-Connes-Charakter

Joachim Cuntz

Received: April 28, 1997

Communicated by Peter Schneider

Abstract. We present a new construction of a bivariant K-functor. The functor can be defined on various categories of topological algebras. The corresponding bivariant theory has a Kasparov product and the other standard properties of KK-theory. We study such a theory in detail on a natural category of locally convex algebras and define a bivariant multiplicative character to bivariant periodic cyclic cohomology.

1991 Mathematics Subject Classification: 18G60, 19K35, 19L10, 46H20, 46L87

Keywords and phrases: bivariant, bivariant K-theory, bivariant Chern character, Chern-Connes-character, locally convex algebra, Frechet algebra, extension, K-theory for topological algebras, cyclic homology for topological algebras

Das Fundament der Nichtkommutativen Geometrie wird gebildet einerseits von Kasparovs KK-Theorie und andererseits von der zyklischen Homologie/Kohomologie von Connes und Tsygan. Diese Theorien verallgemeinern und erweitern zwei wichtige klassische Homologie/Kohomologie-Theorien - nämlich die Atiyah-Hirzebruch-K-Theorie und die de Rham Theorie - von Räumen oder Mannigfaltigkeiten (kommutative Algebren) auf geeignete Kategorien von nichtkommutativen Algebren. Das Wort "verallgemeinern" ist hier nicht völlig angebracht, da diese neuen Theorien angewandt auf den klassischen Fall eine ganz andere neuartige Beschreibung und eine erweiterte Form für die K-Theorie und die de Rham-Theorie geben.
Diese so erweiterten Homologie/Kohomologie-Theorien erlauben es im Prinzip, nichtkommutative Algebren (etwa Algebren von Pseudodifferentialoperatoren) genauso zu behandeln wie Räume, bzw. Algebren von Funktionen. Beide Theorien sind darüberhinaus in natürlicher Weise direkt als bivariante Theorien definiert. Dies stellt einen wichtigen Vorteil dar und ist für Berechnungen der Theorie sehr hilfreich.
Ein wunder Punkt der Theorie war allerdings die Tatsache, dass die K-Homologie sowie die KK-Theorie auf der einen Seite, und die zyklische Theorie auf der anderen, auf verschiedenen Kategorien von topologischen Algebren definiert sind, bzw. sinnvolle

Ergebnisse liefern. Der natürliche Definitionsbereich von Kasparovs KK-Theorie besteht aus C*-Algebren, d.h. aus relativ großen Algebren vom Typ "alle stetigen Funktionen auf einem kompakten Raum". Die zyklische Theorie dagegen liefert vernünftige Ergebnisse nur für wesentlich kleinere Algebren, wie z.B. die Fréchetalgebra aller unendlich oft differenzierbaren Funktionen auf einer Mannigfaltigkeit [Cu4]. Schon wegen des verschiedenen Definitionsbereichs konnten beide Theorien daher nur in speziellen Fällen mit Hilfe etwas künstlicher Tricks miteinander verglichen werden und in diesen Situationen ein partieller bivarianter Chern-Connes-Charakter gefunden werden, siehe z.B. [Co1], [Ks], [Wa], [Ni1].

Andererseits ist bekannt, dass beide Theorien auf ihren verschiedenen Definitionsbereichen ganz analoge Eigenschaften haben. Der letzte wesentliche Schritt hierzu wurde durch den Beweis der Ausschneidungseigenschaft der periodischen zyklischen Theorie in [CuQu2] erzielt. Damit war klar, dass im Prinzip eine allgemeine Transformation von einer Version der KK-Theorie in die bivariante zyklische Theorie zu erwarten ist (bivarianter Chern-Connes-Charakter). Rein algebraisch wurde die Konstruktion eines solchen Charakters schon in [CuQu2] auf Grundlage des Ausschneidungsresultats erläutert.

In der vorliegenden Arbeit führen wir nun eine neue bivariante topologische K-Theorie ein, die auf derselben Kategorie von lokalkonvexen Algebren definiert ist, auf der auch die zyklische Homologie/Kohomologie Sinn macht. Wir bezeichnen diese Theorie mit kk. Wir zeigen, dass kk im wesentlichen dieselben abstrakten Eigenschaften wie die KK-Theorie hat und daher auch in derselben Weise zu berechnen ist. Die Eigenschaften sind Homotopieinvarianz, Stabilität und Ausschneidung, wobei allerdings in der Kategorie der m-Algebren jede dieser Eigenschaften in etwas modifizierter Form zu verstehen ist. Ebenso wie KK kann kk als der universelle Funktor mit diesen drei Eigenschaften charakterisiert werden. Angewendet auf die Algebra der unendlich oft differenzierbaren Funktionen auf einer Mannigfaltigkeit gibt die Theorie natürlich die klassische K-Homologie/K-Theorie. Außerdem ergibt $kk(\mathbb{C},\mathfrak{A})$ die übliche K-Theorie von $\mathfrak{A}$, wenn $\mathfrak{A}$ eine Banachalgebra ist (oder wenn $\mathfrak{A}$ eine Fréchetalgebra ist, unter Verwendung der in [Ph] eingeführten K-Theorie von Fréchetalgebren).

Die Existenz und Multiplikativität des bivarianten Chern-Connes-Charakters folgt im geraden Fall direkt aus der Charakterisierung von kk als universeller Funktor mit gewissen Eigenschaften, da die periodische zyklische Theorie HP^* dieselben Eigenschaften besitzt. Im ungeraden Fall ergibt sich die Existenz des Charakters aus der Ausschneidung für HP^*, und die Multiplikativität aus der Verträglichkeit der Randabbildungen in kk und in HP^*. Diese Verträglichkeit wird durch eine ähnliche Rechnung wie in [Ni2] bewiesen. Im wesentlichen muss das Produkt der Randabbildungen in der Toeplitzerweiterung und in der Einhängungserweiterung bestimmt werden.

Wir beschreiben jetzt kurz den Inhalt der Arbeit. Die ersten beiden Abschnitte enthalten einige allgemeine Grundlagen über die Klasse von lokalkonvexen Algebren, mit der wir arbeiten. Wir nennen diese Algebren m-Algebren. Weiter geben wir Beispiele von m-Algebren und Erweiterungen von m-Algebren, die wir später benutzen. Wir verweisen auf [Ph] für eine ausgezeichnete Zusammenstellung weiterer Konstruktionen in dieser Klasse von topologischen Algebren.

Der dritte Abschnitt enthält mit dem Hauptlemma 3.10 die wesentliche neue technische Idee, die zu einer einfachen und mehr (wenn auch nicht vollständig) algebraischen Konstruktion des Kasparovprodukts führt. Sie erlaubt es, das Produkt ohne die üblichen analytischen Hilfsmittel aus der Theorie der C*-Algebren zu definieren. Übrigens kann die hier eingeführte Strategie auch verwendet werden, um die gewöhnliche KK-Theorie für C*-Algebren oder entsprechende bivariante Theorien für σ-C*-Algebren (siehe [We]) oder Banachalgebren einzuführen. In der Tat gibt unsere Methode ein allgemeines Rezept, um die bivariante K-Theorie mit verschiedenen Homotopieinvarianz- und Stabilitätseigenschaften für verschiedene Kategorien von topologischen Algebren zu konstruieren, siehe Bemerkung 4.6. Sie basiert, ähnlich wie in [Ze] auf Erweiterungen von topologischen Algebren beliebiger Länge und ihren klassifizierenden Abbildungen. Dadurch, dass wir Erweiterungen höherer Länge zulassen, bekommen wir eine einfache Beschreibung des Produkts und vermeiden gleichzeitig eine bekannte Summierbarkeitsobstruktion für "glatte" Erweiterungen der Länge 1, [DoVo].

Abschnitt 4 enthält die Definition und eine Aufstellung der einfachsten Eigenschaften der bivarianten kk-Theorie. Wie in Abschnitt 8 bemerkt wird, ist diese Definition formal verblüffend analog zur Beschreibung der periodischen bivarianten zyklischen Kohomologie, die in [CuQu2, 3.2] enthalten ist. Ein Unterschied zu den üblichen Definitionen der K-Theorie ist, dass wir mit differenzierbaren statt mit stetigen Homotopien arbeiten. Dies ist für die Existenz des Chern-Connes-Chrakters und für die Ausschneidung in kk wichtig. In Abschnitt 5 wird gezeigt, dass jede Erweiterung von m-Algebren, die einen stetigen linearen Schnitt besitzt, lange exakte Folgen in beiden Variablen von kk induziert. Der Beweis benutzt die Methode von [CuSk].

In Abschnitt 6 beweisen wir die Charakterisierung von kk als universeller Funktor, konstruieren den Chern-Connes-Charakter und untersuchen seine Eigenschaften. Insbesondere wird eine Fortsetzung des Charakters auf "p-summierbare" Moduln angegeben, die für Anwendungen und zum Vergleich mit den von Connes und Nistor gegebenen Formeln wichtig ist. Als Nebenprodukt ergibt sich übrigens eine Bestimmung der (stetigen) periodischen zyklischen Homologie/Kohomologie der Schattenideale ℓ^p.

In Abschnitt 7 wird gezeigt, dass $kk_*(\mathbb{C},\mathfrak{A})$ für eine Fréchetalgebra $\mathfrak{A}$ mit der von Phillips definierten K-Theorie $K_*(\mathfrak{A})$ übereinstimmt. Dies ist selbst für $\mathfrak{A} = \mathbb{C}$ a priori überhaupt nicht klar (die kk-Gruppen könnten trivial oder riesengroß sein). Der Beweis benutzt wieder das Hauptlemma 3.10. Wir zeigen auch unabhängig von Phillips' Methoden, dass für Banachalgebren und für gewisse dichte Unteralgebren von Banachalgebren ebenfalls $kk_*(\mathbb{C},\mathfrak{A}) = K_*(\mathfrak{A})$ gilt. Man erhält daher insbesondere eine neue Definition der K-Theorie für die sehr große Klasse der m-Algebren durch

$$K_*(\mathfrak{A}) \underset{def}{=} kk_*(\mathbb{C},\mathfrak{A})$$

Abschnitt 8 enthält einige abschließende Bemerkungen zu der natürlichen Filtrierung auf kk.

Wir erwähnen schließlich, dass das oben beschriebene Dilemma der verschiedenen Definitionsbereiche der KK-Theorie und der zyklischen Theorie prinzipiell auch auf andere Weise gelöst werden kann. Es lässt sich nämlich eine zyklische Theorie entwickeln, die auch für C*-Algebren Sinn macht. Dies wurde im wesentlichen von Puschnigg in [Pu] mit der "asymptotische" zyklischen Theorie auf der Basis eines Vorschlags von Connes-Moscivici [CoMo] erreicht. Die asymptotische Theorie ist aber ihrer Natur nach weniger algebraisch.

Anwendungen der im vorliegenden Artikel dargestellten Theorie bleiben weiteren Arbeiten vorbehalten.

1 m-ALGEBREN UND DIFFERENZIERBARE HOMOTOPIEN

Eine $m-$Algebra ist eine Algebra $\mathfrak{A}$ über $\mathbb{C}$ mit einer vollständigen lokalkonvexen Topologie, die durch eine Familie $\{p_\alpha\}$ von *submultiplikativen* Halbnormen bestimmt ist. Für jedes α gilt also $p_\alpha(xy) \leq p_\alpha(x)p_\alpha(y)$. Die Algebra $\mathfrak{A}$ ist dann eine topologische Algebra, d.h. die Multiplikation ist stetig. Es ist leicht zu sehen, dass m-Algebren gerade die lokalkonvexen Algebren sind, die als projektive Limiten von Banachalgebren darstellbar sind, vgl. [Mi, 5.1]. In [Cu4] wurde gezeigt, dass sich das Argument für die Ausschneidung aus [CuQu2] auf die topologische zyklische Theorie für m-Algebren überträgt.

Die direkte Summe $\mathfrak{A} \oplus \mathfrak{B}$ von zwei $m-$Algebren ist wieder eine $m-$Algebra mit der Topologie, die durch die Halbnormen der Form $p \oplus q$ mit $(p \oplus q)(x, y) = p(x) + q(y)$ definiert ist, wobei p eine stetige Halbnorm auf $\mathfrak{A}$ und q eine stetige Halbnorm auf $\mathfrak{B}$ ist.

Wir erinnern an die Definition des projektiven Tensorprodukts im Sinn von Grothendieck, [Gr], [T]. Für zwei lokalkonvexe Vektorräume V and W ist die projektive Topologie auf dem Tensorprodukt $V \otimes W$ bestimmt durch die Familie der Halbnormen der Form $p \otimes q$, wo p eine stetige Halbnorm auf V und q eine stetige Halbnorm auf W ist. Hierbei ist $p \otimes q$ definiert durch

$$p \otimes q\,(z) = \inf \Big\{ \sum_{i=1}^{n} p(a_i)q(b_i) \,|\, z = \sum_{i=1}^{n} a_i \otimes b_i, a_i \in V, b_i \in W \Big\}$$

für $z \in V \otimes W$. Wir bezeichnen mit $V \hat{\otimes} W$ die Vervollständigung von $V \otimes W$ bezüglich dieser Familie von Halbnormen. Wenn $\mathfrak{A}$ und $\mathfrak{B}$ m-Algebren sind, so ist auch das projektive Tensorprodukt $\mathfrak{A} \hat{\otimes} \mathfrak{B}$ wieder eine m-Algebra (wenn p und q submultiplikativ sind, so auch $p \otimes q$).

Wir geben jetzt einige Beispiele von m-Algebren, die wir später benutzen werden.

1.1 ALGEBREN VON DIFFERENZIERBAREN FUNKTIONEN

Sei $[a, b]$ ein Intervall in $\mathbb{R}$. Wir bezeichnen mit $\mathbb{C}[a, b]$ die Algebra der komplexwertigen $\mathcal{C}^\infty$-Funktionen f auf $[a, b]$, deren Ableitungen in den Endpunkten a und b alle verschwinden (während die 0-te Ableitung, d.h. f selbst, in a und b beliebige Werte annehmen kann).

Eine wichtige Rolle werden auch die Unteralgebren $\mathbb{C}(a, b], \mathbb{C}[a, b)$ and $\mathbb{C}(a, b)$ von $\mathbb{C}[a, b]$ spielen, die nach Definition aus den Funktionen f bestehen, die außerdem noch in a, bzw. in b, bzw. in a und b verschwinden.

Die Topologie auf diesen Algebren ist die übliche Fréchettopologie, die durch die folgende Familie von submultiplikativen Normen p_n definiert ist:

$$p_n(f) = ||f|| + ||f'|| + \tfrac{1}{2}||f''|| + \ldots + \tfrac{1}{n!}||f^{(n)}||$$

Hierbei ist natürlich $||g|| = \sup\{|g(t)| \,\big|\, t \in [a, b]\}$.

Wir bemerken, dass $\mathbb{C}[a, b]$ nuklear im Sinn von Grothendieck [Gr] ist und dass für jeden vollständigen lokalkonvexen Raum V der Raum $\mathbb{C}[a, b] \hat{\otimes} V$ isomorph zu dem

Raum der $\mathcal{C}^\infty$-Funktionen auf $[a,b]$ mit Werten in V ist, deren Ableitungen in beiden Endpunkten verschwinden, [T,§ 51].

Wenn $\mathfrak{A}$ eine m-Algebra ist, schreiben wir $\mathfrak{A}[a,b], \mathfrak{A}[a,b)$ und $\mathfrak{A}(a,b)$ für die m-Algebren $\mathfrak{A}\hat{\otimes}\mathbb{C}[a,b], \mathfrak{A}\hat{\otimes}\mathbb{C}[a,b)$ und $\mathfrak{A}\hat{\otimes}\mathbb{C}(a,b)$.

Zwei stetige lineare Abbildungen $\alpha,\beta : V \to W$ zwischen zwei vollständigen lokalkonvexen Räumen heißen differenzierbar homotop, oder *diffeotop*, falls eine Familie $\varphi_t : V \to W, t \in [0,1]$ von stetigen linearen Abbildungen existiert, so dass $\varphi_0 = \alpha, \varphi_1 = \beta$ und so dass die Abbildung $t \mapsto \varphi_t(x)$ unendlich oft differenzierbar ist für jedes $x \in V$. Eine andere Formulierung dieser Bedingung ist, dass eine stetige lineare Abbildung $\varphi : V \to \mathcal{C}^\infty([0,1])\hat{\otimes}W$ existiert mit der Eigenschaft, dass $\varphi(x)(0) = \alpha(x), \varphi(x)(1) = \beta(x)$ für jedes $x \in V$.

Sei $h : [0,1] \to [0,1]$ eine monotone und bijektive $\mathcal{C}^\infty$-Abbildung, deren Einschränkung auf $(0,1)$ ein Diffeomorphismus $(0,1) \to (0,1)$ ist und deren Ableitungen in 0 und 1 alle verschwinden. Durch Ersetzung von φ_t durch $\psi_t = \varphi_{h(t)}$ sieht man, dass α and β diffeotop sind genau dann, wenn eine stetige lineare Abbildung $\psi : V \to \mathbb{C}[0,1]\hat{\otimes}W$ existiert, für die gilt $\psi(x)(0) = \alpha(x), \psi(x)(1) = \beta(x)$, $x \in V$. Dies zeigt insbesondere, dass Diffeotopie eine Äquivalenzrelation ist.

1.2 Die Tensoralgebra

Es sei V ein vollständiger lokalkonvexer Raum. Wir definieren die Tensoralgebra TV als die Vervollständigung der algebraischen direkten Summe

$$T_{\mathrm{alg}}V = V \oplus V{\otimes}V \oplus V^{\otimes^3} \oplus \dots$$

im Bezug auf die Familie $\{\hat{p}\}$ von Halbnormen, die auf dieser direkten Summe durch

$$\hat{p} = p \oplus p{\otimes}p \oplus p^{\otimes^3} \oplus \dots$$

gegeben sind , wo p alle stetigen Halbnormen auf V durchläuft. Die Zusammensetzung von Tensoren definiert in der üblichen Weise eine Multiplikation auf $T_{\mathrm{alg}}V$, für die die Halbnormen $\hat{p}$ submultiplikativ sind. Die Vervollständigung TV ist daher eine m-Algebra.

Im einfachsten Fall, wo $V = \mathbb{C}$, ist $T\mathbb{C}$ in natürlicher Weise isomorph zu der Algebra der holomorphen Funktionen auf der komplexen Ebene, die im Punkt 0 verschwinden (unter dem Isomorphismus, der eine Folge (λ_n) in $T_{\mathrm{alg}}\mathbb{C}$ auf die Funktion f mit $f(z) = \sum_{n=1}^\infty \lambda_n z^n$ abbildet). Die Topologie ist gegeben durch die Topologie der uniformen Konvergenz auf kompakten Teilmengen

Wir bezeichnen mit $\sigma : V \to TV$ die Abbildung, die V auf den ersten Summanden in $T_{\mathrm{alg}}V$ abbildet. Diese Abbildung σ hat die folgende universelle Eigenschaft: Es sei $s : V \to \mathfrak{A}$ eine beliebige stetige lineare Abbildung von V in eine m-Algebra $\mathfrak{A}$. Dann existiert ein eindeutig bestimmter Homomorphismus $\tau_s : TV \to \mathfrak{A}$ von m-Algebren mit der Eigenschaft, dass $\tau_s \circ \sigma = s$.

Die Tensoralgebra ist differenzierbar kontrahierbar, d.h. die identische Abbildung von TV ist diffeotop zu 0. Eine differenzierbare Familie $\varphi_t : TV \to TV$, für die $\varphi_0 = 0, \varphi_1 = \mathrm{id}$ gilt, ist gegeben durch $\varphi_t = \tau_{t\sigma}, t \in [0,1]$.

1.3 Das freie Produkt von zwei m-Algebren

Zwei m-Algebren $\mathfrak{A}$ und $\mathfrak{B}$ seien gegeben. Das algebraische freie Produkt (in der nichtunitalen Kategorie) von $\mathfrak{A}$ und $\mathfrak{B}$ ist dann die folgende Algebra

$$\mathfrak{A} *_{\mathrm{alg}} \mathfrak{B} = \mathfrak{A} \oplus \mathfrak{B} \oplus (\mathfrak{A} \otimes \mathfrak{B}) \oplus (\mathfrak{B} \otimes \mathfrak{A}) \oplus (\mathfrak{A} \otimes \mathfrak{B} \otimes \mathfrak{A}) \oplus \ldots$$

Die direkte Summe erstreckt sich über alle Tensorprodukte, wo die Faktoren $\mathfrak{A}$ und $\mathfrak{B}$ jeweils abwechselnd auftreten. Die Multiplikation ist, wie bei der Tensoralgebra, die Zusammensetzung von Tensoren, wobei aber anschließend die Multiplikation $\mathfrak{A} \hat{\otimes} \mathfrak{A} \to \mathfrak{A}$ und $\mathfrak{B} \hat{\otimes} \mathfrak{B} \to \mathfrak{B}$ benutzt wird, um alle Terme zu vereinfachen, in denen zwei Elemente in $\mathfrak{A}$ oder zwei Elemente in $\mathfrak{B}$ zusammentreffen.

Wir bezeichnen mit $\mathfrak{A} * \mathfrak{B}$ die Vervollständigung von $\mathfrak{A} *_{\mathrm{alg}} \mathfrak{B}$ bezüglich aller Halbnormen der Form $p * q$ die in der folgenden Weise definiert sind:

$$p * q = p \oplus q \oplus (p \otimes q) \oplus (q \otimes p) \oplus (p \otimes q \otimes p) \oplus \ldots$$

Wir setzen hier alle stetigen Halbnormen p und q auf $\mathfrak{A}$ und $\mathfrak{B}$ ein. Wenn p und q submultiplikativ sind, so ist auch die Halbnorm $p * q$ submultiplikativ und $\mathfrak{A} * \mathfrak{B}$ ist daher eine m-Algebra.

Die Algebra $\mathfrak{A} * \mathfrak{B}$ ist das freie Produkt von $\mathfrak{A}$ und $\mathfrak{B}$ in der Kategorie der m-Algebren. Die kanonischen Inklusionen $\iota_1 : \mathfrak{A} \to \mathfrak{A} * \mathfrak{B}$ und $\iota_2 : \mathfrak{B} \to \mathfrak{A} * \mathfrak{B}$ haben die folgende universelle Eigenschaft: Seien $\alpha : \mathfrak{A} \to \mathfrak{C}$ und $\beta : \mathfrak{B} \to \mathfrak{C}$ zwei stetige Homomorphismen in eine m-Algebra $\mathfrak{C}$. Dann existiert ein eindeutig bestimmter stetiger Homomorphismus $\alpha * \beta : \mathfrak{A} * \mathfrak{B} \to \mathfrak{C}$, so dass $(\alpha * \beta) \circ \iota_1 = \alpha$ und $(\alpha * \beta) \circ \iota_2 = \beta$.

1.4 Die Algebra der glatten kompakten Operatoren

Die Algebra $\mathfrak{K}$ der glatten kompakten Operatoren besteht aus allen Matrizen (a_{ij}) mit schnell abfallenden Matrixelementen $a_{ij} \in \mathbb{C}$, $i, j = 0, 1, 2\ldots$ (für eine andere Beschreibung dieser Algebra siehe [ENN]). Die Topologie auf $\mathfrak{K}$ ist gegeben durch die Familie von Normen p_n, $n = 0, 1, 2\ldots$, die durch

$$p_n\left((a_{ij})\right) = \sum_{i,j} |1 + i + j|^n \, |a_{ij}|$$

definiert sind. Man prüft leicht nach, dass die p_n submultiplikativ sind und dass $\mathfrak{K}$ vollständig ist. Damit ist $\mathfrak{K}$ eine m-Algebra. Als linearer lokalkonvexer Raum ist $\mathfrak{K}$ natürlich isomorph zum Folgenraum s und daher nuklear.

Die Abbildung, die $(a_{ij}) \otimes (b_{kl})$ auf die $\mathbb{N}^2 \times \mathbb{N}^2$-Matrix $(a_{ij} b_{kl})_{(i,k)(j,l) \in \mathbb{N}^2 \times \mathbb{N}^2}$ abbildet, gibt offensichtlich einen Isomorphismus Θ zwischen $\mathfrak{K} \hat{\otimes} \mathfrak{K}$ und $\mathfrak{K}$ (vgl. auch [Ph,2.7])

LEMMA 1.4.1 *Sei $\Theta : \mathfrak{K} \to \mathfrak{K} \hat{\otimes} \mathfrak{K}$ die oben angegebene Abbildung und $\iota : \mathfrak{K} \to \mathfrak{K} \hat{\otimes} \mathfrak{K}$ die Inklusionsabbildung, die x auf $e^{00} \otimes x$ abbildet (wo e^{00} die Matrix mit Elementen a_{ij} ist, für die $a_{ij} = 1$, falls $i = j = 0$, und $a_{ij} = 0$ sonst). Dann ist Θ diffeotop zu ι. Dasselbe gilt für die entsprechenden Abbildungen $\Theta' : \mathfrak{K} \to M_2(\mathfrak{K})$ und $\iota' : \mathfrak{K} \to M_2(\mathfrak{K})$.*

Beweis: Wir können $\mathfrak{K}$ darstellen als eine Algebra von Operatoren auf dem Raum s der schnell fallenden Folgen. Die gesuchte Homotopie kann durch direkte Summen von Rotationen in jeweils zweidimensionalen Teilräumen, die Vektoren der Form $\xi_i \otimes \xi_j$ in der Standardbasis von $s\hat{\otimes}s$ in Vektoren der Form $\xi_0 \otimes \xi_{ij}$ überführen, realisiert werden. Dabei bezeichnet ξ_{ij} eine Umnumerierung der Basis von s mit Indexmenge $\mathbb{N} \times \mathbb{N}$; siehe auch [Ph,2.7]. *q.e.d.*

BEMERKUNG 1.4.2 *Sei V ein Banachraum. Dann besteht $\mathfrak{K}\hat{\otimes}V$ gerade aus den Matrizen, oder den durch $\mathbb{N} \times \mathbb{N}$ indizierten Folgen $(v_{ij})_{i,j\in\mathbb{N}}$, für die der Ausdruck*

$$\bar{p}_n((v_{ij})) \underset{def}{=} \sum_{i,j} (1+i+j)^n \|v_{ij}\|$$

endlich ist für jedes n. Die Topologie auf $\mathfrak{K}\hat{\otimes}V$ ist natürlich gerade durch die Normen $\bar{p}_n$ gegeben. Um dies zu sehen, betrachten wir das Tensorprodukt α_n der Norm p_n auf $\mathfrak{K}$ mit der auf V gegebenen Norm $\|\cdot\|$. Wenn dann x^{ij} die Matrix bezeichnet, die $x \in V$ als i,j-tes Element hat und sonst 0 ist, so gilt

$$\alpha_n(x^{ij}) = (1+i+j)\|x\|$$

nach [T, Prop. 43.1]. Dies zeigt sofort, dass

$$\alpha_n((v_{ij})) \leq \bar{p}_n((v_{ij}))$$

für alle Matrizen (v_{ij}) im algebraischen Tensorprodukt $\mathfrak{K}\otimes V$. Die umgekehrte Ungleichung folgt aus der Definition der projektiven Tensornorm. Daher ist für jedes feste n die Vervollständigung $(\mathfrak{K}\otimes V)_{\bar{p}_n}$ isometrisch isomorph zu $(\mathfrak{K})_{p_n}\hat{\otimes}V$ und besteht gerade aus den Matrizen (v_{ij}), für die $\bar{p}_n((v_{ij}))$ endlich ist.

1.5 DIE GLATTE TOEPLITZALGEBRA

Die Elemente der Algebra $\mathcal{C}^\infty S^1$ können als Potenzreihen in dem Erzeuger z (definiert durch $z(t) = t$, $t \in S^1 \subset \mathbb{C}$) geschrieben werden. Die Koeffizienten sind schnell abfallend, d.h. genauer gilt

$$\mathcal{C}^\infty(S^1) = \Big\{ \sum_{k\in\mathbb{Z}} a_k z^k \;\mid\; \sum_{k\in\mathbb{Z}} |a_k|\,|k|^n < \infty \text{ für jedes feste } n \in \mathbb{N} \Big\}$$

Submultiplikative Normen, die die Topologie beschreiben, sind gegeben durch

$$q_n\left(\sum a_k z^k\right) = \sum |1+k|^n\,|a_k|$$

Als topologischer Vektorraum ist die glatte Toeplitzalgebra $\mathfrak{T}$ dann definiert als die direkte Summe $\mathfrak{T} = \mathfrak{K} \oplus \mathcal{C}^\infty(S^1)$.

Um die Multiplikation in $\mathfrak{T}$ zu definieren, schreiben wir v_k für das Element $(0, z^k)$ von $\mathfrak{T}$ und einfach x für das Element $(x, 0)$ mit $x \in \mathfrak{K}$. Außerdem bezeichnet e^{ij} das Element von $\mathfrak{T}$, das durch die Matrix (a_{kl}) mit $a_{kl} = 1$, falls $k = i, l = j$, und $a_{kl} = 0$

sonst, bestimmt ist (mit der Vereinbarung, dass $e^{ij} = 0$, wenn $i < 0$ oder $j < 0$). Die Multiplikation in $\mathfrak{T}$ ist dann bestimmt durch die folgenden Regeln:

$$e^{ij}e^{kl} = \delta_{jk}e^{il}$$

$$v_k e^{ij} = e^{(i+k),j} \qquad e^{ij}v_k = e^{i,(j-k)}$$

$(i,j,k \in \mathbb{Z})$; und

$$v_k v_{-l} = \begin{cases} v_{k-l}(1 - E_{l-1}) & l > 0 \\ v_{k-l} & l \leq 0 \end{cases}$$

wo $E_l = e^{00} + e^{11} + \ldots + e^{ll}$. Wenn p_n die in 1.4 definierten Normen auf $\mathfrak{K}$ sind und q_n die oben definierten Normen auf $\mathcal{C}^\infty(S^1)$, so ist leicht zu sehen, dass jede Norm der Form $p_n \oplus q_m$ submultiplikativ auf $\mathfrak{T} = \mathfrak{K} \oplus \mathcal{C}^\infty(S^1)$ mit der so definierten Multiplikation ist. Es ist offensichtlich, dass $\mathfrak{K}$ ein abgeschlossenes Ideal in $\mathfrak{T}$ ist, und dass der Quotient $\mathfrak{T}/\mathfrak{K}$ gerade $\mathcal{C}^\infty(S^1)$ ist.

1.6 ABGELEITETE UNTERALGEBREN VON BANACHALGEBREN

Viele der wichtigsten m-Algebren sind von einem speziellen Typ - sie sind Algebren von "nichtkommutativen $\mathcal{C}^\infty$-Funktionen". Um diese Klasse von Fréchetalgebren zu charakterisieren, verwenden wir die Ideen aus [BlCu], wo der Fall von abgeleiteten Unteralgebren von C*-Algebren eingehend untersucht wurde.
Sei A eine Banachalgebra. Eine abgeleitete Unteralgebra von A ist eine Unteralgebra $\mathfrak{A}$, für die gilt

1) Auf $\mathfrak{A}$ ist eine Familie $p_0, p_1, \ldots$ von Halbnormen gegeben, wo p_0 ein Vielfaches der gegebenen Norm auf A ist. $\mathfrak{A}$ ist vollständig im Bezug auf diese Familie.

2) Für jedes k gilt

$$p_k(xy) \leq \sum_{i+j=k} p_i(x)p_j(y), \qquad x, y \in \mathfrak{A}$$

Falls 1) und 2) erfüllt sind, so ist für jedes k die Summe $p_0 + p_1 + \ldots + p_k$ eine submultiplikative Norm. $\mathfrak{A}$ ist daher gleichzeitig eine Fréchetalgebra und eine m-Algebra. Eines der wichtigsten Beispiele ist $\mathcal{C}^\infty[0,1]$ mit den Halbnormen $p_n(f) = \frac{1}{n!}\|f^{(n)}\|$ oder allgemeiner $\mathcal{C}^\infty M$ für eine differenzierbare kompakte Mannigfaltigkeit M.
Wir erinnern daran, dass eine Unteralgebra $\mathfrak{A}$ einer Banachalgebra A *abgeschlossen unter holomorphem Funktionalkalkül* ist, falls das Spektrum $Sp(x)$ jedes Elements x von $\mathfrak{A}$, in A und $\mathfrak{A}$ dasselbe ist und falls außerdem für jede in einer Umgebung von $Sp(x)$ holomorphe Funktion f, auch $f(x)$ wieder in $\mathfrak{A}$ liegt.

LEMMA 1.6.1 *Wenn $\mathfrak{A} \subset A$ die Bedingungen 1) und 2) erfüllt, so ist $\mathfrak{A}$ abgeschlossen unter holomorphem Funktionalkalkül.*

Beweis: vgl. [BlCu, 3.12 oder 6.4]. Sei A_k die Vervollständigung von $\mathfrak{A}$ bezüglich der Norm $\|\cdot\|_k = p_0 + p_1 + \ldots p_k$. Für alle $x, y \in \mathfrak{A}$ gilt

$$\|xy\|_{k+1} \leq \|x\|_k\|y\|_{k+1} + \|x\|_{k+1}\|y\|_k$$

Dies impliziert, dass

$$\limsup \sqrt[n]{\|x^{2n}\|_{k+1}} \leq \limsup \sqrt[n]{\|x^n\|_k} \limsup \sqrt[n]{\|x^n\|_{k+1}}$$

für jedes $x \in \mathfrak{A}$ und damit für die Spektralradien

$$r_{A_{k+1}}(x)^2 = r_{A_{k+1}}(x^2) \leq r_{A_k}(x) r_{A_{k+1}}(x)$$

und somit, dass $r_{A_{k+1}}(x) = r_{A_k}(x)$.
Dasselbe Argument gilt für die Algebren $\widetilde{A_k}$, wo noch eine Eins adjungiert wurde.
Falls nun $x \in \widetilde{\mathfrak{A}}$ invertierbar in $\widetilde{A}$ ist, so existiert $\varepsilon \geq 0$, so dass für jedes $y \in \mathfrak{A}$ mit $\|x^{-1} - y\| \leq \varepsilon$ gilt, dass $r_A(1 - xy) < 1$. Daher ist $r_{A_k}(1 - xy) < 1$ für alle k und somit xy und also auch x invertierbar in A_k (nach einem Diagonalfolgenargument ist $\mathfrak{A}$ der Durchschnitt aller Bilder von A_k in A).
Dies zeigt, dass $\mathrm{Sp}_{\mathfrak{A}} x = \mathrm{Sp}_A x$ für alle $x \in \mathfrak{A}$. Wenn jetzt f eine Funktion ist, die holomorph in einer Umgebung von $\mathrm{Sp}_{\mathfrak{A}} x = \mathrm{Sp}_A x$ ist, so liegt $f(x)$ in A_k für alle k und damit auch in $\mathfrak{A}$. *q.e.d.*

BEMERKUNG 1.6.2 *Falls $\mathfrak{A}$ eine abgeleitete Unteralgebra einer C^*-Algebra ist, so ist $\mathfrak{A}$ sogar invariant unter Funktionalkalkül mit $\mathcal{C}^\infty$-Funktionen, siehe [BlCu, 6.4].*

LEMMA 1.6.3 *Seien $\mathfrak{A}$ und $\mathfrak{B}$ abgeleitete Unteralgebren von A bzw. B. Dann ist $\mathfrak{A} \hat{\otimes} \mathfrak{B}$ eine abgeleitete Unteralgebra von $A \hat{\otimes} B$.*

Beweis: Falls $p_0, p_1, \ldots$ und $q_0, q_1, \ldots$ die Familien von Halbnormen mit der Eigenschaft 2) sind, die die Topologien auf $\mathfrak{A}$ und $\mathfrak{B}$ bestimmen, so ist $u_0, u_1, \ldots$ mit

$$u_k = \sum_{i+j=k} p_i \otimes q_j$$

eine Familie von Halbnormen auf $\mathfrak{A} \hat{\otimes} \mathfrak{B}$, für die $\mathfrak{A} \hat{\otimes} \mathfrak{B}$ vollständig ist und für die 2) gilt. *q.e.d.*

Wir bezeichnen mit $\mathcal{K}_1$ die Banachalgebra der komplexen Matrizen $(a_{ij})_{i,j \in \mathbb{N}}$ mit

$$\|(a_{ij})\|_1 = \sum |a_{ij}| < \infty$$

LEMMA 1.6.4 *$\mathfrak{K}$ ist eine abgeleitete Unteralgebra von $\mathcal{K}_1$.*

Beweis: Die Topologie von $\mathfrak{K}$ ist bestimmt durch die Halbnormen $\alpha_0, \alpha_1, \alpha_2 \ldots$ mit

$$\alpha_n((a_{ij})) = \frac{1}{n!} \sum (i+j)^n \|a_{ij}\|$$

Nach Definition ist $\alpha_0 = \| \cdot \|_1$. Die Gleichung

$$\frac{1}{n!}(i+j)^n = \sum_{r+s=n} \frac{1}{r!} i^r \frac{1}{s!} j^s$$

zeigt, dass $\alpha_n(xy) \leq \sum_{r+s=n} \alpha_r(x) \alpha_s(y)$. *q.e.d.*

LEMMA 1.6.5 *Sei $\alpha : A \to B$ ein stetiger Homomorphismus zwischen Banachalgebren und $\mathfrak{B} \subset B$ eine abgeleitete Unteralgebra mit definierendem System von Halbnormen $q_0, q_1, \ldots$.*
Dann ist $\mathfrak{A} = \alpha^{-1}(\mathfrak{B})$ mit dem System $p_0, p_1, \ldots$ von Halbnormen, wo

$$p_0 = C\|\cdot\|_A \qquad C = \max(1, \|\alpha\|)$$
$$p_i(x) = q_i(\alpha(x)),\ i = 1, 2, \ldots$$

eine abgeleitete Unteralgebra von A.

Beweis: Klar. *q.e.d.*

2 EINIGE WICHTIGE ERWEITERUNGEN VON m-ALGEBREN

In der bivarianten K-Theorie für C*-Algebren spielen eine Reihe von Standarderweiterungen eine grundlegende Rolle. Wir beschreiben in diesem Abschnitt zunächst einmal die analogen Erweiterungen in der Kategorie der m-Algebren. Hierbei ist zu beachten, dass außerdem jeweils Algebren von stetigen Funktionen durch die entsprechenden Algebren von $\mathcal{C}^\infty$-Funktionen ersetzt werden, da wir statt mit stetigen Homotopien mit differenzierbaren Homotopien arbeiten werden. Darüberhinaus benötigen wir aber auch noch weitere Erweiterungen, die bisher in der K-Theorie noch nicht so stark in Erscheinung getreten sind. Insbesondere wird die universelle Erweiterung durch die Tensoralgebra in unserer Theorie eine tragende Rolle spielen.

Wir betrachten in erster Linie Erweiterungen, die stetige lineare Schnitte besitzen, d.h. als exakte Folgen von lokalkonvexen Vektorräumen einfach direkte Summen darstellen. Wir nennen solche Erweiterungen linear zerfallend. Das Tensorprodukt einer linear zerfallenden Erweiterung mit einer beliebigen lokalkonvexen Algebra ist wieder linear zerfallend.

Die meisten Erweiterungen in diesem Abschnitt sind außerdem von dem Typ, dass die Algebra in der Mitte kontrahierbar ist, so dass die Ideale verschiedene Formen der Einhängung (des Quotienten) beschreiben.

2.1 DIE EINHÄNGUNGSERWEITERUNG.

Dies ist das Analogon zu der fundamentalen Erweiterung der algebraischen Topologie. Sie hat die folgende Form

$$0 \to \mathbb{C}(0,1) \to \mathbb{C}[0,1) \to \mathbb{C} \to 0$$

oder allgemeiner

$$0 \to \mathfrak{A}(0,1) \to \mathfrak{A}[0,1) \to \mathfrak{A} \to 0$$

mit einer beliebigen m-Algebra $\mathfrak{A}$.

Wir erinnern daran, dass $\mathbb{C}(0,1)$ und $\mathbb{C}[0,1)$ Algebren von $\mathcal{C}^\infty$- Funktionen auf dem Intervall $[0,1]$, deren Ableitungen alle in 0 und 1 verschwinden, bezeichnen, und dass die Algebra $\mathbb{C}[0,1)$ differenzierbar kontrahierbar ist, vgl. 1.1.

2.2 DIE UNIVERSELLE ERWEITERUNG.

Auf dieser Erweiterung beruht unsere Definition der bivarianten K-Theorie für m-Algebren. Für eine m-Algebra $\mathfrak{A}$ ist die Tensoralgebra $T\mathfrak{A}$ über dem lokalkonvexen Raum $\mathfrak{A}$ wie in Abschnitt 1 definiert. Wenn wir die Tatsache verwenden, dass $\mathfrak{A}$ auch eine Algebra ist und die universelle Eigenschaft von $T\mathfrak{A}$ auf die Abbildung id:$\mathfrak{A} \to \mathfrak{A}$ anwenden, so erhalten wir einen Homomorphismus $\alpha = \tau_{\mathrm{id}} : T\mathfrak{A} \to \mathfrak{A}$ (ein Element $x_1 \otimes x_2 \otimes \ldots \otimes x_n$ von $T\mathfrak{A}$ wird dabei auf $x_1 x_2 \ldots x_n$ in $\mathfrak{A}$) abgebildet. Wir definieren jetzt $J\mathfrak{A}$ als den Kern von α. Die Erweiterung

$$0 \to J\mathfrak{A} \to T\mathfrak{A} \stackrel{\alpha}{\to} \mathfrak{A} \to 0$$

besitzt dann einen stetigen linearen Schnitt. Die m-Algebra $T\mathfrak{A}$ ist glatt kontrahierbar. Die universelle Eigenschaft dieser Erweiterung wird im nächsten Abschnitt erläutert und benutzt werden.

2.3 DIE GLATTE TOEPLITZERWEITERUNG.

Die glatte Toeplitzalgebra $\mathfrak{T}$ wurde in 1.5 eingeführt. Nach Konstruktion enthält $\mathfrak{T}$ die Algebra $\mathfrak{K}$ als Ideal und wir erhalten die folgende Erweiterung

$$0 \to \mathfrak{K} \to \mathfrak{T} \stackrel{\pi}{\to} \mathcal{C}^\infty(S^1) \to 0$$

die natürlich nach Konstruktion auch einen stetigen linearen Schnitt erlaubt.

Sei nun $\kappa : \mathfrak{T} \to \mathbb{C}$ der kanonische Homomorphismus, der v_1 und v_{-1} auf 1 abbildet und $\mathfrak{T}_0 = \operatorname{Ker} \kappa$. Durch Restriktion der Toeplitzerweiterung erhalten wir die folgende Erweiterung

$$0 \to \mathfrak{K} \to \mathfrak{T}_0 \to \mathcal{C}_0^\infty(S^1 \backslash 1) \to 0$$

Wir werden später sehen, dass $\mathfrak{T}_0$ "kk-kontrahierbar" ist.

2.4 DIE UNIVERSELLE ZWEIFACH TRIVIALE ERWEITERUNG.

Mit einer m-Algebra $\mathfrak{A}$ assoziieren wir wie in [Cu2] die Algebra $Q\mathfrak{A} = \mathfrak{A} * \mathfrak{A}$. Wir bezeichnen mit ι und $\bar{\iota}$ die beiden kanonischen Inklusionen von $\mathfrak{A}$ in $Q\mathfrak{A}$. Die Algebra $Q\mathfrak{A}$ ist in natürlicher Weise $\mathbb{Z}/2$-graduiert durch den involutiven Automorphismus τ, der $\iota(\mathfrak{A})$ und $\bar{\iota}(\mathfrak{A})$ vertauscht.

Das Ideal $q\mathfrak{A}$ in $Q\mathfrak{A}$ ist definiert als der Kern des kanonischen Homomorphismus $\pi = \mathrm{id} * \mathrm{id} : \mathfrak{A} * \mathfrak{A} \to \mathfrak{A}$. Die Erweiterung

$$0 \to q\mathfrak{A} \to Q\mathfrak{A} \xrightarrow{\pi} \mathfrak{A} \to 0 \tag{1}$$

besitzt dann zwei verschiedene Schnitte, die Algebrenhomomorphismen sind; nämlich ι und $\bar{\iota}$. Sie hat die folgende universelle Eigenschaft: Sei

$$0 \to \mathfrak{E}_0 \to \mathfrak{E}_1 \to \mathfrak{A} \to 0 \tag{2}$$

eine Erweiterung mit zwei verschiedenen Schnitten $\alpha, \bar{\alpha} : \mathfrak{A} \to \mathfrak{E}_1$, die stetige Algebrahomomorphismen sind. Dann existiert ein Morphismus (d.h. ein kommutatives

Diagramm von Abbildungen) von der Erweiterung (1) in die Erweiterung (2) wie folgt:

$$\begin{array}{ccccccc} 0 \to & q\mathfrak{A} & \to & Q\mathfrak{A} & \to & \mathfrak{A} & \to 0 \\ & \downarrow {\scriptstyle \alpha * \bar\alpha} & & \downarrow {\scriptstyle \alpha * \bar\alpha} & & \downarrow {\scriptstyle \mathrm{id}} & \\ 0 \to & \mathfrak{E}_0 & \to & \mathfrak{E}_1 & \to & \mathfrak{A} & \to 0 \end{array}$$

Dieser Morphismus führt nach Konstruktion die Schnitte ι und $\bar\iota$ in α und $\bar\alpha$ über.

2.5 Die Erweiterung, die die gerade und die ungerade K-Theorie verbindet.

Wir konstruieren in diesem Artikel die bivariante K-Theorie aus Erweiterungen, d.h. wir benutzen das "ungerade" oder Ext-Bild. Die folgende Erweiterung erlaubt es, diesen Zugang mit dem "gerade" Bild von [Cu2] zu vergleichen. Sie wird in Abschnitt 7 eine wichtige Rolle spielen. Wie oben seien $\iota, \bar\iota : \mathfrak{A} \to Q\mathfrak{A}$ die kanonischen Inklusionen. Wir setzen

$$\mathfrak{E} := \{ f \in Q\mathfrak{A}[0,1] \mid \exists x \in \mathfrak{A}, f(0) = \iota(x), f(1) = \bar\iota(x),\ f(t) - f(0) \in q\mathfrak{A}, t \in [0,1] \}$$

Die Erweiterung

$$0 \to q\mathfrak{A}(0,1) \to \mathfrak{E} \to \mathfrak{A} \to 0$$

besitzt dann einen stetigen linearen Schnitt, der $x \in \mathfrak{A}$ auf $f \in \mathfrak{E}$ mit $f(t) = (1-t)\iota(x) + t\bar\iota(x)$ abbildet.

3 Morphismen von der universellen Erweiterung.

Als erstes analysieren wir die universelle Eigenschaft der Erweiterung $0 \to J\mathfrak{A} \to T\mathfrak{A} \to \mathfrak{A} \to 0$ aus 2.2.

Satz 3.1 *Es sei*

$$0 \to \mathfrak{E}_0 \to \mathfrak{E}_1 \underset{\curvearrowleft s}{\xrightarrow{\pi}} \mathfrak{A} \to 0$$

eine Erweiterung mit einem stetigen linearen Schnitt s (d.h. $\pi s = \mathrm{id}_{\mathfrak{A}}$). Weiter sei $\varphi : \mathfrak{A}' \to \mathfrak{A}$ ein Homomorphismus und $\tau_{s\varphi} : T\mathfrak{A}' \to \mathfrak{E}_1$ der Homomorphismus, der sich wie in 1.2 aus der universellen Eigenschaft der Tensoralgebra $T\mathfrak{A}'$ ergibt. Dann existiert ein eindeutig bestimmter Homomorphismus $\gamma_{s\varphi} : J\mathfrak{A}' \to \mathfrak{E}_0$ so dass das folgende Diagramm kommutiert

$$\begin{array}{ccccccc} 0 \to & \mathfrak{E}_0 & \to & \mathfrak{E}_1 & \overset{\pi}{\to} & \mathfrak{A} \to 0 \\ & \uparrow \gamma_{s\varphi} & & \uparrow \tau_{s\varphi} & & \uparrow \varphi \\ 0 \to & J\mathfrak{A}' & \to & T\mathfrak{A}' & \to & \mathfrak{A}' \to 0 \end{array}$$

Beweis: Das Bild von $J\mathfrak{A}'$ unter $\tau_{s\varphi}$ ist in $\mathfrak{E}_0$ enthalten, weil die Abbildung $\pi \circ \tau_{s\varphi}$ das Ideal $J\mathfrak{A}'$ annulliert und weil andererseits $\mathfrak{E}_0 = \operatorname{Ker} \pi$. Wir setzen $\gamma_{s\varphi} = \tau_{s\varphi}|_{J\mathfrak{A}'}$ *q.e.d.*

Die Anwendung dieses Prinzips auf die in Abschnitt 2 eingeführten Erweiterungen ergibt Homomorphismen $J\mathfrak{A} \to \mathfrak{A}(0,1), J(\mathcal{C}^\infty(S^1)) \to \mathfrak{K}$ und $J\mathfrak{A} \to q\mathfrak{A}(0,1)$, die im folgenden immer wieder benutzt werden.
Wenn man das Resultat auf die Erweiterung $0 \to J\mathfrak{A} \to T\mathfrak{A} \to \mathfrak{A} \to 0$ anwendet, sieht man insbesondere, dass $\mathfrak{A} \mapsto J\mathfrak{A}$ ein Funktor ist: Jeder Homomorphismus $\varphi : \mathfrak{A}' \to \mathfrak{A}$ induziert einen Homomorphismus $J\mathfrak{A}' \to J\mathfrak{A}$, den wir mit $J(\varphi)$ bezeichnen.

LEMMA 3.2 *Sei* $0 \to \mathfrak{E}_0 \to \mathfrak{E}_1 \overset{\overset{s}{\curvearrowleft}}{\xrightarrow{\pi}} \mathfrak{A} \to 0$ *eine Erweiterung mit stetigem linearen Schnitt* s *und* $\varphi : \mathfrak{A}' \to \mathfrak{A}$ *ein Homomorphismus wie in 3.1.*

(a) Sei s' *ein weiterer stetiger linearer Schnitt. Dann ist* $\gamma_{s'\varphi} : J\mathfrak{A} \to \mathfrak{E}_0$ *diffeotop zu* $\gamma_{s\varphi}$.

(b) Wenn ein stetiger linearer Schnitt s'' *existiert, der ein Algebrenhomomorphismus ist, so ist* $\gamma_{s\varphi}$ *diffeotop zu* 0.

(c) Wenn ein Algebrenhomomorphismus $\varphi' : \mathfrak{A}' \to \mathfrak{E}_1$ *existiert mit* $\pi \circ \varphi' = \varphi$, *so ist* $\gamma_{s\varphi}$ *diffeotop zu* 0.

Beweis: (a) Setze $s_t = ts' + (1-t)s$. Dann ist $\gamma_{s_t}, t \in [0,1]$ eine differenzierbare Homotopie, die γ_s und $\gamma_{s'}$ verbindet. (b) und (c) folgen aus (a) und aus der Tatsache, dass die Einschränkungen von $\tau_{s''\varphi}$ und $\tau_{\varphi'}$ auf $J\mathfrak{A}'$ verschwinden. *q.e.d.*

Für $\varphi = \mathrm{id}$ nennen wir γ_s die klassifizierende Abbildung zu der linear zerfallenden Erweiterung

$$0 \to \mathfrak{E}_0 \to \mathfrak{E}_1 \overset{\overset{s}{\curvearrowleft}}{\xrightarrow{\pi}} \mathfrak{A} \to 0$$

Das nächste einfache Lemma beschreibt das Verhalten der klassifizierenden Abbildung unter Morphismen (d.h. kommutativen Diagrammen) von Erweiterungen. Es wird in den folgenden Abschnitten implizit immer wieder benutzt.

LEMMA 3.3 *Betrachte das folgende kommutative Diagramm von Erweiterungen*

$$\begin{array}{ccccccc}
0 \to & \mathfrak{E}_0 & \to & \mathfrak{E}_1 & \to & \mathfrak{A} \to 0 \\
 & \uparrow\psi_0 & & \uparrow\psi_1 & & \uparrow\varphi \\
0 \to & \mathfrak{E}'_0 & \to & \mathfrak{E}'_1 & \to & \mathfrak{A}' \to 0
\end{array}$$

mit stetigen linearen Schnitten $s : \mathfrak{A} \to \mathfrak{E}_1$ *und* $s' : \mathfrak{A}' \to \mathfrak{E}'_1$.
Es gilt $\gamma_{s\varphi} = \gamma_s \circ J(\varphi)$ *und diese Abbildung ist diffeotop zu* $\psi_0 \circ \gamma_{s'}$ *(falls* $s\varphi = \psi_1 s'$, *so gilt sogar* $\gamma_s \circ J(\varphi) = \psi_0 \circ \gamma_{s'}$ *).*

DEFINITION-SATZ 3.4 *Gegeben seien zwei Erweiterungen von* $\mathfrak{A}$

$$0 \to \mathfrak{E}_0 \to \mathfrak{E}_1 \overset{\overset{s}{\curvearrowleft}}{\xrightarrow{\pi}} \mathfrak{A} \to 0$$

$$0 \to \mathfrak{E}_0 \to \mathfrak{E}'_1 \overset{\overset{s'}{\curvearrowleft}}{\xrightarrow{\pi'}} \mathfrak{A} \to 0$$

mit stetigen linearen Schnitten. Die Summe dieser beiden Erweiterungen ist nach Definition die Erweiterung

$$0 \to M_2(\mathfrak{E}_0) \to \mathcal{D} \to \mathfrak{A} \to 0$$

wo $\mathcal{D} = \left\{ \begin{pmatrix} x & a \\ b & x' \end{pmatrix} \mid x \in \mathfrak{E}_1, x' \in \mathfrak{E}'_1, \pi(x) = \pi'(x'); a, b \in \mathfrak{E}_0 \right\}$.
Sie erlaubt $s \oplus s' = \begin{pmatrix} s & 0 \\ 0 & s' \end{pmatrix}$ *als stetigen linearen Schnitt. Der assoziierte Homomorphismus* $\gamma_{s\oplus s'} : J\mathfrak{A} \to M_2(\mathfrak{E}_0)$ *ist gegeben durch*

$$\gamma_{s\oplus s'} = \gamma_s \oplus \gamma_{s'} = \begin{pmatrix} \gamma_s & 0 \\ 0 & \gamma_{s'} \end{pmatrix}$$

Beweis: Klar. *q.e.d.*

Als Beispiel betrachten wir die glatte Toeplitzerweiterung

$$0 \to \mathfrak{K} \to \mathfrak{T} \xrightarrow{\pi} \mathcal{C}^\infty(S^1) \to 0 \tag{3}$$

aus 2.3. Es sei u der Automorphismus von $\mathcal{C}^\infty(S^1)$, der die Orientierung von S^1 umkehrt. Dann ist die Summe von (3) mit der Erweiterung

$$0 \to \mathfrak{K} \to \mathfrak{T} \xrightarrow{u\pi} \mathcal{C}^\infty(S^1) \to 0 \tag{4}$$

trivial (d.h. sie erlaubt einen stetigen linearen Schnitt, der ein Algebrenhomomorphismus ist). In der Tat ist die Abbildung, die die k-te Potenz z^k des Erzeugers z von $\mathcal{C}^\infty(S^1)$, $k \in \mathbb{Z}$ auf die k-te Potenz der Matrix $\begin{pmatrix} v_1 & e^{00} \\ 0 & v_{-1} \end{pmatrix}$ (mit den Bezeichnungen von 1.5) abbildet, ein stetiger Homomorphismus. Wenn daher s der stetige lineare Schnitt $\mathcal{C}^\infty(S^1) \to \mathfrak{T}$ ist, der z^k auf v_k abbildet und s' der Schnitt für (4) der z^k auf v_{-k} abbildet, so ist $\gamma_s \oplus \gamma_{s'}$ diffeotop zu 0.

DEFINITION-SATZ 3.5 *Gegeben seien m-Algebren* $\mathfrak{A}$ *und* $\mathfrak{B}$. *Wenn* $\varphi : \mathfrak{A} \to \mathfrak{B}$ *ein Homomorphismus zwischen m-Algebren ist, so bezeichnen wir mit* $\langle\varphi\rangle$ *die Äquivalenzklasse von* φ *im Bezug auf die Relation der Diffeotopie und wir setzen*

$$\langle \mathfrak{A}, \mathfrak{B} \rangle = \{ \langle\varphi\rangle \mid \varphi \text{ ist ein stetiger Homomorphismus } \mathfrak{A} \to \mathfrak{B} \}$$

Für Homomorphismen $\alpha, \beta : \mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{B}$ *definieren wir wie in 3.4 die direkte Summe* $\alpha \oplus \beta$ *als*

$$\alpha \oplus \beta = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} : \mathfrak{A} \longrightarrow M_2(\mathfrak{K}\hat{\otimes}\mathfrak{B}) \cong \mathfrak{K}\hat{\otimes}\mathfrak{B}$$

Mit der durch $\langle\alpha\rangle + \langle\beta\rangle = \langle\alpha \oplus \beta\rangle$ *definierten Addition ist die Menge* $\langle \mathfrak{A}, \mathfrak{K}\hat{\otimes}\mathfrak{B}\rangle$ *der Diffeotopieklassen von Homomorphismen von* $\mathfrak{A}$ *nach* $\mathfrak{K}\hat{\otimes}\mathfrak{B}$ *eine abelsche Halbgruppe mit Nullelement* $\langle 0 \rangle$.

Beweis: Dies folgt aus Lemma 1.4.1. *q.e.d.*

Für jede m-Algebra $\mathfrak{A}$ ist $J\mathfrak{A}$ wieder eine m-Algebra. Wir können daher durch Iteration $J^2\mathfrak{A} = J(J\mathfrak{A}), \ldots, J^n\mathfrak{A} = J(J^{n-1}\mathfrak{A})$ bilden. Abbildungen von $J^n\mathfrak{A}$ in eine m-Algebra $\mathfrak{B}$ gehören dann zu Erweiterungen der Länge n.

DEFINITION-SATZ 3.6 *Eine exakte Folge*

$$0 \longrightarrow \mathfrak{E}_0 \xrightarrow{\varphi_0} \mathfrak{E}_1 \xrightarrow{\varphi_1} \ldots \longrightarrow \mathfrak{E}_n \xrightarrow{\varphi_n} \mathfrak{A} \longrightarrow 0$$

wo $\mathfrak{E}_0, \ldots, \mathfrak{E}_n, \mathfrak{A}$ m-Algebren und die φ_i stetige Homomorphismen sind, heiße linear zerfallende n-Schritt-Erweiterung, falls sie als exakte Folge von lokalkonvexen Vektorräumen zerfällt (d.h. falls $\mathfrak{E}_i \cong \operatorname{Ker}\varphi_i \oplus \operatorname{Im}\varphi_{i-1}$). Jede Wahl $s_1, \ldots, s_n$ von stetigen linearen Schnitten (d.h. $\varphi_i s_i$ ist für alle i eine stetige Projektion auf $\operatorname{Im}\varphi_i$) bestimmt in eindeutiger Weise einen Homomorphismus $\gamma_{(s_1,\ldots,s_n)} : J^n\mathfrak{A} \to \mathfrak{E}_0$ und Homomorphismen $\gamma_{(s_{k+1},\ldots,s_n)} : J^{n-k}\mathfrak{A} \to \mathfrak{E}_k$ so dass das folgende Diagramm kommutiert

$$\begin{array}{ccccccccc}
0 \longrightarrow & \mathfrak{E}_0 \xrightarrow{\varphi_0} & \mathfrak{E}_1 & \xrightarrow{\varphi_1} \ldots & \mathfrak{E}_{n-1} \xrightarrow{\varphi_{n-1}} & \mathfrak{E}_n \xrightarrow{\varphi_n} & \mathfrak{A} & \longrightarrow 0 \\
 & \uparrow \gamma_{(s_1\ldots,s_n)} & \uparrow \tau_{s_1}\gamma_{(s_2\ldots,s_n)} & & \uparrow \tau_{s_{n-1}}\gamma_{s_n} & \uparrow \tau_{s_n} & \| & \\
0 \longrightarrow & J^n\mathfrak{A} \longrightarrow & TJ^{n-1}\mathfrak{A} & \to \ldots & TJ\mathfrak{A} \longrightarrow & T\mathfrak{A} \longrightarrow & \mathfrak{A} & \longrightarrow 0
\end{array}$$

Wenn $s_1', \ldots, s_n'$ eine andere Familie von stetigen linearen Schnitten ist, so ist $\gamma_{(s_1',\ldots,s_n')}$ diffeotop zu $\gamma_{(s_1,\ldots,s_n)}$.

In dem vorhergehenden Diagramm interessieren wir uns in erster Linie für die klassifizierende Abbildung $\gamma = \gamma_{(s_1,\ldots,s_n)}$. Diese hängt bis auf Diffeotopie nicht von $(s_1, \ldots, s_n)$ sondern nur von der gegebenen n-Schritt-Erweiterung ab.

Betrachten wir zwei Erweiterungen der Länge n und der Länge m

$$0 \longrightarrow \mathfrak{E}_0 \xrightarrow{\varphi_0} \mathfrak{E}_1 \xrightarrow{\varphi_1} \ldots \longrightarrow \mathfrak{E}_n \xrightarrow{\varphi_n} \mathfrak{A} \longrightarrow 0 \tag{5}$$

und

$$0 \longrightarrow \mathfrak{E}_0' \xrightarrow{\varphi_0'} \mathfrak{E}_1' \xrightarrow{\varphi_1'} \ldots \longrightarrow \mathfrak{E}_m' \xrightarrow{\varphi_m'} \mathfrak{A}' \longrightarrow 0 \tag{6}$$

wo $\mathfrak{E}_0' = \mathfrak{A}$. Das wohlbekannte Yonedaprodukt besteht in der Zusammensetzung dieser zwei Erweiterungen zu einer Erweiterung der Länge $n+m$ von der Form

$$0 \longrightarrow \mathfrak{E}_0 \xrightarrow{\varphi_0} \mathfrak{E}_1 \xrightarrow{\varphi_1} \ldots \longrightarrow \mathfrak{E}_n \xrightarrow{\varphi_0'\varphi_n} \mathfrak{E}_1' \xrightarrow{\varphi_1'} \ldots \longrightarrow \mathfrak{E}_m' \xrightarrow{\varphi_m'} \mathfrak{A}' \longrightarrow 0 \tag{7}$$

LEMMA 3.7 *Es seien $\gamma : J^n\mathfrak{A} \longrightarrow \mathfrak{E}_0$ und $\gamma' : J^m\mathfrak{A}' \longrightarrow \mathfrak{E}_0' = \mathfrak{A}$ die Abbildungen, die mit (5) und (6) assoziiert sind. Die klassifizierende Abbildung $J^{n+m}\mathfrak{A}' \longrightarrow \mathfrak{E}_0$ zu der Erweiterung (7) ist gegeben durch $\gamma \circ J^n(\gamma')$.*

Beweis: Dies folgt aus 3.3. *q.e.d.*

DEFINITION 3.8 *Es sei $\varphi : J\mathfrak{A} \to \mathcal{C}^\infty(S^1)\hat{\otimes}\mathfrak{A}$ die Komposition der klassifizierenden Abbildung $J\mathfrak{A} \to \mathfrak{A}(0,1)$ zu der Erweiterung*

$$0 \longrightarrow \mathfrak{A}(0,1) \longrightarrow \mathfrak{A}[0,1) \longrightarrow \mathfrak{A} \longrightarrow 0$$

mit der Inklusionsabbildung $\mathfrak{A}(0,1) \longrightarrow \mathcal{C}^\infty(S^1)\hat{\otimes}\mathfrak{A}$. *Wir bezeichnen mit* ε *die Abbildung*

$$\varepsilon : J^2\mathfrak{A} \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{A}$$

die unter Benutzung von φ *zu der Erweiterung*

$$0 \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{A} \longrightarrow \mathfrak{T}\hat{\otimes}\mathfrak{A} \overset{s}{\underset{\curvearrowleft}{\longrightarrow}} \mathcal{C}^\infty(S^1)\hat{\otimes}\mathfrak{A} \longrightarrow 0$$

gehört (d.h. $\varepsilon = \gamma_{s\varphi}$*).*

Man beachte, dass eine linear zerfallende Erweiterung in der Kategorie der lokalkonvexen Vektorräume einfach eine direkte Summe darstellt und daher natürlich auch nach Tensorieren mit beliebigen lokalkonvexen Räumen noch exakt bleibt.

Durch Hintereinanderschaltung der Abbildungen $J^4\mathfrak{A} \xrightarrow{J^2(\varepsilon)} J^2(\mathfrak{K}\hat{\otimes}\mathfrak{A})$, sowie $J^2(\mathfrak{K}\hat{\otimes}\mathfrak{A}) \to \mathfrak{K}\hat{\otimes}J^2(\mathfrak{A})$ und $\mathfrak{K}\hat{\otimes}J^2\mathfrak{A} \xrightarrow{\mathrm{id}\otimes\varepsilon} \mathfrak{K}\hat{\otimes}\mathfrak{K}\hat{\otimes}\mathfrak{A}$ bekommmen wir, unter leichtem Missbrauch der Bezeichnungen,

$$\varepsilon^2 : J^4\mathfrak{A} \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{K}\hat{\otimes}\mathfrak{A} \cong \mathfrak{K}\hat{\otimes}\mathfrak{A}$$

und, nach Induktion

$$\varepsilon^n : J^{2n}\mathfrak{A} \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{A}$$

Wir können bei der Konstruktion von ε statt der Toeplitzerweiterung auch die inverse Toeplitzerweiterung verwenden und erhalten dann eine Abbildung $\varepsilon_- : J^2\mathfrak{A} \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{A}$, die nach 3.4 die Eigenschaft hat, dass $\varepsilon \oplus \varepsilon_-$ diffeotop zu 0 ist.

LEMMA 3.9 *Für jedes Paar von m-Algebren* $\mathfrak{A}$ *und* $\mathfrak{B}$ *existieren kanonische Abbildungen* $J(\mathfrak{A}\hat{\otimes}\mathfrak{B}) \to J\mathfrak{A}\hat{\otimes}\mathfrak{B}$ *und* $J(\mathfrak{A}\hat{\otimes}\mathfrak{B}) \to \mathfrak{A}\hat{\otimes}J\mathfrak{B}$, *die mit den folgenden linear zerfallenden Erweiterungen assoziiert sind*

$$\begin{array}{ccccccccc} 0 & \to & J\mathfrak{A}\hat{\otimes}\mathfrak{B} & \to & T\mathfrak{A}\hat{\otimes}\mathfrak{B} & \to & \mathfrak{A}\hat{\otimes}\mathfrak{B} & \to & 0 \\ 0 & \to & \mathfrak{A}\hat{\otimes}J\mathfrak{B} & \to & \mathfrak{A}\hat{\otimes}T\mathfrak{B} & \to & \mathfrak{A}\hat{\otimes}\mathfrak{B} & \to & 0\,. \end{array}$$

Wir bemerken, dass insbesondere für jede m-Algebra $\mathfrak{A}$ ein kanonischer Homomorphismus $J(\mathfrak{A}) \to J(\mathbb{C})\hat{\otimes}\mathfrak{A}$ existiert. Es ist klar, dass die in 3.8 definierte Abbildung $\varepsilon = \varepsilon_{\mathfrak{A}} : J^2\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ als Komposition der Abbildung $J^2\mathfrak{A} \to J^2\mathbb{C}\hat{\otimes}\mathfrak{A}$ mit der Abbildung $\varepsilon \otimes \mathrm{id}_{\mathfrak{A}}$ geschrieben werden kann.

Das folgende Lemma bildet den Kernpunkt für unsere Konstruktion des Produkts der in Abschnitt 4 definierten bivarianten K-Theorie.

HAUPTLEMMA 3.10 $\mathfrak{A}$ *und* $\mathfrak{B}$ *seien m-Algebren und* γ_+, γ_- *die zwei Abbildungen von* $J^2(\mathfrak{A}\hat{\otimes}\mathfrak{B})$ *nach* $J\mathfrak{A}\hat{\otimes}J\mathfrak{B}$, *die sich durch Anwendung von 3.9, wie folgt in den zwei möglichen Weisen ergeben:*

$$\begin{array}{ccccc} & & J(J\mathfrak{A}\hat{\otimes}\mathfrak{B}) & & \\ & \nearrow & & \searrow & \\ J^2(\mathfrak{A}\hat{\otimes}\mathfrak{B}) & & \overset{\gamma_+}{\longrightarrow} \atop \underset{\gamma_-}{\longrightarrow} & & J\mathfrak{A}\hat{\otimes}J\mathfrak{B} \\ & \searrow & & \nearrow & \\ & & J(\mathfrak{A}\hat{\otimes}J\mathfrak{B}) & & \end{array}$$

Dann ist die Abbildung $\gamma_+ \oplus \gamma_- = \begin{pmatrix} \gamma_+ & 0 \\ 0 & \gamma_- \end{pmatrix} : J^2(\mathfrak{A}\hat{\otimes}\mathfrak{B}) \to M_2(J\mathfrak{A}\hat{\otimes}J\mathfrak{B})$ *diffeotop zu 0.*

Beweis: Betrachte die folgende Erweiterung

$$0 \longrightarrow J\mathfrak{A}\hat{\otimes}J\mathfrak{B} \longrightarrow T\mathfrak{A}\hat{\otimes}J\mathfrak{B} + J\mathfrak{A}\hat{\otimes}T\mathfrak{B} \underset{\pi}{\overset{s}{\rightleftarrows}} \mathfrak{A}\hat{\otimes}J\mathfrak{B} \oplus J\mathfrak{A}\hat{\otimes}\mathfrak{B} \longrightarrow 0$$

Die Algebra in der Mitte wird hier als Unteralgebra von $T\mathfrak{A}\hat{\otimes}T\mathfrak{B}$ angesehen. Die Abbildung $\gamma_+ \oplus \gamma_-$ ist durch Rotationen in 2×2-Matrizen diffeotop zu $\gamma_{s\alpha}$, wenn $\alpha : J(\mathfrak{A}\hat{\otimes}\mathfrak{B}) \to \mathfrak{A}\hat{\otimes}J\mathfrak{B} \oplus J\mathfrak{A}\hat{\otimes}\mathfrak{B}$ die natürliche Abbildung bezeichnet. Zum Beweis der Behauptung genügt es daher nach Lemma 3.2 (c) zu zeigen, dass ein Homomorphismus $\alpha' : J(\mathfrak{A}\hat{\otimes}\mathfrak{B}) \longrightarrow T\mathfrak{A}\hat{\otimes}J\mathfrak{B} + J\mathfrak{A}\hat{\otimes}T\mathfrak{B}$ existiert, für den $\pi \circ \alpha' = \alpha$ gilt.

Nun kann aber α' als klassifizierende Abbildung $\gamma_{s'}$ in der linear zerfallenden Erweiterung

$$0 \longrightarrow J\mathfrak{A}\hat{\otimes}T\mathfrak{B} + T\mathfrak{A}\hat{\otimes}J\mathfrak{B} \longrightarrow T\mathfrak{A}\hat{\otimes}T\mathfrak{B} \overset{s'}{\rightleftarrows} \mathfrak{A}\hat{\otimes}\mathfrak{B} \longrightarrow 0$$

gewählt werden. Die Tatsache, dass $\pi \circ \gamma_{s'} = \alpha$ folgt aus den zwei folgenden kommutativen Diagrammen

$$\begin{array}{ccccccc}
0 \longrightarrow & J\mathfrak{A}\hat{\otimes}T\mathfrak{B} + T\mathfrak{A}\hat{\otimes}J\mathfrak{B} & \longrightarrow & T\mathfrak{A}\hat{\otimes}T\mathfrak{B} & \longrightarrow & \mathfrak{A}\hat{\otimes}\mathfrak{B} & \longrightarrow 0 \\
 & \downarrow & & \downarrow & & \| & \\
0 \longrightarrow & \mathfrak{A}\hat{\otimes}J\mathfrak{B} & \longrightarrow & \mathfrak{A}\hat{\otimes}T\mathfrak{B} & \longrightarrow & \mathfrak{A}\hat{\otimes}\mathfrak{B} & \longrightarrow 0
\end{array}$$

und

$$\begin{array}{ccccccc}
0 \longrightarrow & J\mathfrak{A}\hat{\otimes}T\mathfrak{B} + T\mathfrak{A}\hat{\otimes}J\mathfrak{B} & \longrightarrow & T\mathfrak{A}\hat{\otimes}T\mathfrak{B} & \longrightarrow & \mathfrak{A}\hat{\otimes}\mathfrak{B} & \longrightarrow 0 \\
 & \downarrow & & \downarrow & & \| & \\
0 \longrightarrow & J\mathfrak{A}\hat{\otimes}\mathfrak{B} & \longrightarrow & T\mathfrak{A}\hat{\otimes}\mathfrak{B} & \longrightarrow & \mathfrak{A}\hat{\otimes}\mathfrak{B} & \longrightarrow 0
\end{array}$$

sowie aus Lemma 3.3. *q.e.d.*

Als nächstes soll die Abbildung $\varepsilon : J^k\mathfrak{A} \longrightarrow \mathfrak{K}\hat{\otimes}J^{k-2}\mathfrak{A}$, die in 3.8 eingeführt wurde, genauer untersucht werden. Zur besseren Übersichtlichkeit schreiben wir J_i für die i-te Anwendung des J-Funktors. D.h. also $J^k\mathfrak{A} = J_k J_{k-1} \dots J_1(\mathfrak{A})$.

Für jede Wahl von i, j mit $1 \le j < i \le k$, ergibt die Anwendung von 3.9 eine Abbildung $\eta_{ij} : J^k\mathfrak{A} \longrightarrow J^2\mathbb{C}\hat{\otimes}J^{k-2}\mathfrak{A}$, indem wir das j-te und das i-te J im Tensorprodukt $\mathbb{C}\hat{\otimes}\mathfrak{A}$ auf $\mathbb{C}$ und alle anderen J auf den zweiten Faktor $\mathfrak{A}$ anwenden. Explizit sieht also η_{ij} folgendermaßen aus:

$$\eta_{ij} : J_k \dots J_1(\mathfrak{A}) \longrightarrow J_i J_j(\mathbb{C})\hat{\otimes}J_k \dots \check{J}_i \dots \check{J}_j \dots J_1(\mathfrak{A})$$

wo $\vee$ Auslassung bedeutet.

Wenn wir dies mit der Abbildung $\varepsilon : J^2\mathbb{C} = J_i J_j(\mathbb{C}) \longrightarrow \mathfrak{K}$ kombinieren, erhalten wir eine Familie von Abbildungen $\varepsilon_{ij} : J^k\mathfrak{A} \longrightarrow \mathfrak{K}\hat{\otimes}J^{k-2}\mathfrak{A}$, $1 \le j < i \le k$. (Mit dieser Bezeichnungsweise wäre die unter 3.8 betrachtete Abbildung ε_{21}).

KOROLLAR 3.11 *Es gelten die folgenden differenzierbaren Homotopien*

(a) $\varepsilon_{i-1,j} \oplus \varepsilon_{i,j} \sim 0, \quad 1 < j < i-1 \leq k-1$

$\varepsilon_{i,j-1} \oplus \varepsilon_{i,j} \sim 0, \quad 2 \leq j < i \leq k$

(b) Für alle $i, j,\ 1 \leq i, j \leq k-1$, *gilt* $\varepsilon_{i+1,i} \sim \varepsilon_{j+1,j}$

Hierbei bezeichnet $\sim$ *Diffeotopie.*

Beweis: (a) ergibt sich aus 3.10. (b) folgt aus (a) unter Benutzung der Tatsache, dass die Menge der Diffeotopieklassen von Homomorphismen $J^k\mathfrak{A}$ nach $\mathfrak{K}\hat{\otimes}J^{k-2}\mathfrak{A}$ nach 3.5 eine abelsche Halbgruppe mit 0-Element ist. In dieser Halbgruppe sind die Klassen von $\varepsilon_{i+1,i}$ und von $\varepsilon_{i,i-1}$ beide invers zu $\varepsilon_{i+1,i-1}$, und daher gleich. *q.e.d.*

4 DER BIVARIANTE K-FUNKTOR

Wir sind jetzt soweit, dass wir das eigentliche Untersuchungsobjekt dieser Arbeit einführen können. Wir betrachten die Menge der Diffeotopieklassen von Homomorphismen $H_k = \langle J^k\mathfrak{A}, \mathfrak{K}\hat{\otimes}\mathfrak{B}\rangle$, wobei $H_0 = \langle\mathfrak{A}, \mathfrak{K}\hat{\otimes}\mathfrak{B}\rangle$. Jedes H_k ist eine abelsche Halbgruppe mit der üblichen K-Theorie-Addition $\langle\alpha\rangle + \langle\beta\rangle = \langle\alpha\oplus\beta\rangle$, siehe 3.6. Die Klasse $\langle 0\rangle$ ist das Nullelement.

Es existiert eine kanonische Abbildung $S : H_k \longrightarrow H_{k+2}$, die man in der folgenden Weise erhält: für $\langle\alpha\rangle \in H_k, \alpha : J^k\mathfrak{A} \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{B}$, sei $S\langle\alpha\rangle = \langle(\mathrm{id}_{\mathfrak{K}} \otimes \alpha)\circ\varepsilon\rangle$. Dabei ist $\varepsilon : J^{k+2}\mathfrak{A} \longrightarrow \mathfrak{K}\otimes J^k\mathfrak{A}$ die in 3.8 betrachtete Abbildung (genauer gesagt $\varepsilon = \varepsilon_{k+2,k+1}$ mit den Bezeichnungen von 3.9). Weiter sei $\varepsilon_- : J^{k+2}\mathfrak{A} \longrightarrow \mathfrak{K}\otimes J^k\mathfrak{A}$ die Abbildung, die sich in derselben Weise, aber unter Ersetzung der Toeplitzerweiterung durch die inverse Toeplitzerweiterung, ergibt. Die Diskussion nach 3.4 zeigt, dass die Summe $\varepsilon\oplus\varepsilon_-$ diffeotop zu 0 ist. Daher ist $S\langle\alpha\rangle + S_-\langle\alpha\rangle = 0$, wenn wir $S_-\langle\alpha\rangle = \langle(\mathrm{id}_{\mathfrak{K}}\otimes\alpha)\circ\varepsilon_-\rangle$ setzen.

DEFINITION 4.1 *Es seien* $\mathfrak{A}$ *und* $\mathfrak{B}$ *m-Algebren und* $* = 0$ *oder* 1. *Wir setzen*

$$kk_*(\mathfrak{A}, \mathfrak{B}) = \varinjlim_k H_{2k+*} = \varinjlim_k \langle J^{2k+*}\mathfrak{A}, \mathfrak{K}\hat{\otimes}\mathfrak{B}\rangle$$

Die vorhergehende Diskussion zeigt, dass $kk_*(\mathfrak{A}, \mathfrak{B})$ nicht nur eine abelsche Halbgruppe, sondern sogar eine abelsche Gruppe ist (jedes Element besitzt ein Inverses).

Die wesentliche Eigenschaft von kk_* ist das Produkt, das mit Hilfe des Hauptlemmas 3.10 definiert werden kann. Wir benötigen für die Definition noch einige Bezeichnungen.

Wenn $\alpha : J^k\mathfrak{A} \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{B}$ ein Homomorphismus ist, so bezeichne α_j den Homomorphismus $\alpha_j : J^{k+j}\mathfrak{A} \to \mathfrak{K}\hat{\otimes}J^j\mathfrak{B}$, der durch Hintereinanderschaltung von $J^j(\alpha) : J^{k+j}\mathfrak{A} \to J^j(\mathfrak{K}\hat{\otimes}\mathfrak{B})$ mit der kanonischen Abbildung $J^j(\mathfrak{K}\hat{\otimes}\mathfrak{B}) \to \mathfrak{K}\hat{\otimes}J^j\mathfrak{B}$ entsteht; cf. 3.9.

LEMMA 4.2 *Mit den Bezeichnungen vom Ende des Abschnitt 3 sind die folgenden Abbildungen* $J^{k+j+2}\mathfrak{A} \to \mathfrak{K}\hat{\otimes}J^j\mathfrak{B}$ *diffeotop* $(\sim)$

$$(a) \quad ((id_{\mathfrak{K}}\otimes\alpha)\circ\varepsilon_{k+2,k+1})_j \sim (id_{\mathfrak{K}}\otimes\alpha_j)\circ\varepsilon_{k+j+2,k+j+1}$$

$$(b) \quad (id_{\mathfrak{K}}\otimes\alpha_j)\circ\varepsilon_{k+j+2,k+j+1} \sim (id_{\mathfrak{K}}\otimes\varepsilon_{j+2,j+1})\circ\alpha_{j+2}$$

Beweis: (a) ist eine Konsequenz von Korollar 3.11 und (b) folgt sofort aus Lemma 1.4.1. *q.e.d.*

THEOREM 4.3 *(a) Es existiert ein assoziatives und in beiden Variablen additives Produkt*

$$kk_i(\mathfrak{A},\mathfrak{B}) \times kk_j(\mathfrak{B},\mathfrak{C}) \longrightarrow kk_{i+j}(\mathfrak{A},\mathfrak{C})$$

($i,j \in \mathbb{Z}/2; \mathfrak{A},\mathfrak{B}$ und $\mathfrak{C}$ m-Algebren), das für $\alpha : J^n\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{B}\,, \beta : J^m\mathfrak{B} \to \mathfrak{K}\hat{\otimes}\mathfrak{C}$ in der folgenden Weise definiert ist:

$$\langle\alpha\rangle \cdot \langle\beta\rangle = \langle(id_{\mathfrak{K}} \otimes \beta) \circ \alpha_m\rangle$$

(b) Es existiert ein bilineares graduiert kommutatives äußeres Produkt

$$kk_i(\mathfrak{A}_1,\mathfrak{A}_2) \times kk_j(\mathfrak{B}_1,\mathfrak{B}_2) \longrightarrow kk_{i+j}(\mathfrak{A}_1\hat{\otimes}\mathfrak{A}_2,\mathfrak{B}_1\hat{\otimes}\mathfrak{B}_2)$$

Beweis: (a) Die einzige Behauptung, die nicht offensichtlich ist, ist die, dass das Produkt wohldefiniert ist. Wir müssen zeigen, dass unsere Definition des Produkts verträglich ist mit den Identifikationen in dem induktiven Limes, der in der Definition von kk_* in 4.1 benutzt wird. Dafür müssen wir nachprüfen, dass

$$\beta \circ (\alpha \circ \varepsilon)_j \sim (\beta \circ \alpha_j) \circ \varepsilon$$
$$(\beta \circ \varepsilon) \circ \alpha_{j+2} \sim (\beta \circ \alpha_j) \circ \varepsilon$$

(Wir haben hier bei den Bezeichnungen die Indizes von ε, die nach 3.11 irrelevant sind, und das Tensorprodukt mit $\mathrm{id}_{\mathfrak{K}}$ weggelassen.) Die Existenz dieser Diffeotopien ist genau die Aussage von Lemma 4.2.

(b) Dies folgt sofort aus der Existenz der natürlichen Abbildungen

$$J^{2n+2m+i+j}(\mathfrak{A}_1\hat{\otimes}\mathfrak{A}_2) \longrightarrow (J^{2n+i}\mathfrak{A}_1)\hat{\otimes}(J^{2m+j}\mathfrak{A}_2)$$

vgl. 3.9. *q.e.d.*

Lemma 3.7 zeigt, dass das (innere) Produkt in (a) gerade dem Yonedaprodukt von Erweiterungen entspricht.

SATZ 4.4 *kk_* hat die folgenden Eigenschaften*

(a) Jeder Homomorphismus $\varphi : \mathfrak{A} \to \mathfrak{B}$ definiert ein Element $kk(\varphi)$ in der Gruppe $kk_0(\mathfrak{A},\mathfrak{B})$. Wenn $\psi : \mathfrak{B} \to \mathfrak{C}$, ein weiterer Homomorphismus ist, so gilt

$$kk(\psi \circ \varphi) = kk(\varphi) \cdot kk(\psi)$$

$kk_*(\mathfrak{A},\mathfrak{B}\,)$ ist ein kontravarianter Funktor in $\mathfrak{A}$ und ein kovarianter Funktor in $\mathfrak{B}$. Wenn $\alpha : \mathfrak{A}' \to \mathfrak{A}$ und $\beta : \mathfrak{B} \to \mathfrak{B}'$ Homomorphismen sind, so sind die in der ersten und zweiten Variablen von kk_* induzierten Abbildungen gegeben durch Linksmultiplikation mit $kk(\alpha)$ und Rechtsmultiplikation mit $kk(\beta)$.

(b) Für jede m-Algebra $\mathfrak{A}$ ist $kk_*(\mathfrak{A},\mathfrak{A})$ ein $\mathbb{Z}/2$-graduierter Ring mit Einselement $kk(\mathrm{id}_{\mathfrak{A}})$.

(c) Der Funktor kk_* ist invariant unter Diffeotopien in beiden Variablen.

(d) Die kanonische Inklusion $\iota : \mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ definiert ein invertierbares Element in $kk_0(\mathfrak{A}, \mathfrak{K}\hat{\otimes}\mathfrak{A})$. Insbesondere ist $kk_*(\mathfrak{A}, \mathfrak{B}) \cong kk_*(\mathfrak{K}\hat{\otimes}\mathfrak{A}, \mathfrak{B})$ und $kk_*(\mathfrak{B}, \mathfrak{A}) \cong kk_*(\mathfrak{B}, \mathfrak{K}\hat{\otimes}\mathfrak{A})$ für jede m-Algebra $\mathfrak{B}$.

Beweis: (a) Die Diffeotopieklasse $\langle\alpha\rangle$ von α ist ein Element von H_0 und damit nach Definition auch von kk_0. Die zweite Behauptung folgt sofort aus der Definition des Produkts.
(b) Dies folgt aus 4.3. Das Einselement ist $kk(\mathrm{id}_{\mathfrak{A}}) \in kk_0(\mathfrak{A}, \mathfrak{A})$.
(c) Die Abbildungen $\mathfrak{A} \to \mathfrak{A}[0,1]$ und $\mathfrak{A}[0,1] \to \mathfrak{A}$, die a auf $a \cdot 1$ und f auf $f(0)$ abbilden, definieren Elemente in $kk_0(\mathfrak{A}, \mathfrak{A}[0,1])$ und $kk_0(\mathfrak{A}[0,1], \mathfrak{A})$, die invers zueinander sind.
(d) folgt aus Lemma 1.4.1. *q.e.d.*

Nach Definition bestimmt ε ein Element in $kk_0(\mathfrak{A}, \mathfrak{A})$ und zwar dasselbe wie $\mathrm{id}_{\mathfrak{A}}$, d.h. also das Einselement. Andererseits kann die Abbildung $\varepsilon : J^2\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ auch als Element von $kk_0(J^2\mathfrak{A}, \mathfrak{A})$ oder als Element von $kk_1(J\mathfrak{A}, \mathfrak{A})$ gedeutet werden.

SATZ 4.5 *Die Abbildung $\varepsilon : J^2\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ definiert invertierbare Elemente e_0 in $kk_0(J^2\mathfrak{A}, \mathfrak{A})$ und e_1 in $kk_1(J\mathfrak{A}, \mathfrak{A})$.*

Beweis: Die Inversen zu e_0 und e_1 sind gegeben durch $\mathrm{id}_{J^2\mathfrak{A}}$ und $\mathrm{id}_{J\mathfrak{A}}$. *q.e.d.*

Insbesondere ist also

$$kk_1(\mathfrak{A}, \mathfrak{B}) \cong kk_0(J\mathfrak{A}, \mathfrak{B}) \cong kk_0(\mathfrak{A}, J\mathfrak{B})$$

BEMERKUNG 4.6 *Die hier entwickelte Konstruktion der bivarianten K-Theorie ist sehr allgemein und kann ohne weiteres verwendet werden, um bivariante Theorien mit verschiedenen Stabilitäts- und Homotopieinvarianzeigenschaften auch für ganz andere Kategorien von topologischen Algebren einzuführen. Benötigt werden hierzu für jede Algebra $\mathcal{A}$ in einer solchen Kategorie die folgenden Erweiterungen:*

(a) die universelle Erweiterung $0 \to J\mathcal{A} \to T\mathcal{A} \to \mathcal{A} \to 0$

(b) die Einhängungserweiterung $0 \to \mathcal{A}(0,1) \to \mathcal{A}(0,1] \to \mathcal{A} \to 0$

(c) die Toeplitzerweiterung $0 \to \mathcal{K} \otimes \mathcal{A} \to \mathcal{T} \otimes \mathcal{A} \to \mathcal{A}(S^1) \to 0$

Hierbei ist $\otimes$ ein geeignetes Tensorprodukt in der Kategorie, $\mathcal{K}$ eine Vervollständigung der Algebra M_∞ der endlichen Matrizen beliebiger Grösse, sowie $\mathcal{A}(0,1)$, $\mathcal{A}(0,1]$, $\mathcal{A}(S^1)$ geeignete Algebren von Funktionen auf $(0,1)$, $(0,1]$, S^1 mit Werten in $\mathcal{A}$. Die universelle Erweiterung muß universell für eine gewisse Klasse von Erweiterungen sein (bei m-Algebren für linear zerfallende Erweiterungen). Außerdem müssen die Einhängungserweiterung und die Toeplitzerweiterung zusammensetzbar sein, d.h. es muß eine Abbildung $\mathcal{A}(0,1) \to \mathcal{A}(S^1)$ existieren.
Diese Bedingungen sind zum Beispiel erfüllt in der Kategorie der C^*-Algebren mit der üblichen Toeplitzerweiterung und mit der universellen C^*-Algebra- Vervollständigung von $T\mathcal{A}$, für die die kanonische lineare Inklusion $\mathcal{A} \to T\mathcal{A}$ involutionserhaltend und

von Norm ≤ 1 ist. Damit ist die entsprechende Erweiterung universell für Erweiterungen, die einen stetigen Schnitt mit Norm 1 erlauben. Dieselben Wahlen funktionieren in der Kategorie der σ-C^*-Algebren.
In der Kategorie der Banachalgebren kann für $\mathcal{K}$ die Algebra $\mathcal{K}_1$ aus 1.6.4 und für die Funktionenalgebren die einmal stetig differenzierbaren Funktionen mit Werten in $\mathcal{A}$ verwendet werden. Eine geeignete Wahl für das Tensorprodukt ist hier auch das projektive.
Die Stabilitäts- und Homotopieinvarianzeigenschaften der Theorie sind dann bestimmt durch die Wahl der Algebra $\mathcal{K}$ und der Funktionenalgebren (stetige oder differenzierbare Funktionen mit Werten in $\mathcal{A}$). Die Größe von $\mathcal{K}$ korrespondiert aufgrund der Toeplitzerweiterung zur Größe der Funktionenalgebren. Die hier dargestellte Theorie ist gewissermaßen minimal (für die Größe von $\mathcal{K}$ und der Funktionenalgebren) mit der Eigenschaft, dass die oben erwähnte Abbildung $\mathcal{A}(0,1) \to \mathcal{A}(S^1)$ noch existiert.
Wenn wir nur Erweiterungen der Länge 1, d.h. Abbildungen $J\mathcal{A} \to \mathcal{K} \otimes \mathcal{A}$ zulassen würden, so müsste nach der Summierbarkeitsobstruktion von Douglas-Voiculescu [DoVo], die Algebra $\mathcal{K}$ alle Schattenideale ℓ^p für $p \geq 1$ enthalten. Dadurch, dass wir Abbildungen $J^n\mathcal{A} \to \mathcal{K} \otimes \mathcal{A}$ für beliebige n verwenden, erhalten wir das Produkt und umgehen gleichzeitig diese Obstruktion.

5 AUSSCHNEIDUNG UND DIE LANGEN EXAKTEN FOLGEN IN BEIDEN VARIABLEN

In diesem Abschnitt halten wir uns eng an das in [CuSk] gegebene Argument für die Ausschneidung. Ein Unterschied hier ist, dass wir nur differenzierbare Homotopien, d.h. Diffeotopien benutzen. Der Beweisgang zeigt übrigens interessanterweise auch, dass dies wirklich wesentlich ist. Wenn wir kk mit Hilfe von stetigen Homotopien definiert hätten, würde die Ausschneidung nicht gelten; siehe Bemerkung 5.6. Weiter wird ein Teil des Arguments im Vergleich zu [CuSk] dadurch vereinfacht, dass die inverse Bottabbildung $\varepsilon : J^2\mathbb{C} \to \mathfrak{K}$ in unsere Theorie schon eingebaut ist und nach Definition das Einselement von $kk_0(\mathbb{C},\mathbb{C})$ repräsentiert.

Wenn $\alpha : \mathfrak{A} \to \mathfrak{B}$ ein Homomorphismus zwischen m-Algebren ist, werden wir im folgenden mit $\mathfrak{K}(\alpha), \alpha(0,1), \alpha[0,1], J(\alpha)\ldots$ die induzierten Abbildungen $\mathfrak{K}\hat{\otimes}\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{B}$, $\mathfrak{A}(0,1) \to \mathfrak{B}(0,1)$, $\mathfrak{A}[0,1) \to \mathfrak{B}[0,1)$, $J\mathfrak{A} \to J\mathfrak{B}\ldots$ bezeichnen.

Wie üblich definieren wir auch den (differenzierbaren) Abbildungskegel C_α durch

$$C_\alpha = \{(x,f) \in \mathfrak{A} \oplus \mathfrak{B}[0,1) \mid \alpha(x) = f(0)\}$$

LEMMA 5.1 *Sei $\mathfrak{D}$ eine m-Algebra und $\alpha : \mathfrak{A} \to \mathfrak{B}$ ein Homomorphismus*

(a) Die Folge

$$kk_*(\mathfrak{D}, C_\alpha) \xrightarrow{\cdot kk(\pi)} kk_*(\mathfrak{D},\mathfrak{A}) \xrightarrow{\cdot kk(\alpha)} kk_*(\mathfrak{D},\mathfrak{B})$$

ist exakt. Hierbei bezeichnet $\pi : C_\alpha \to \mathfrak{A}$ die Projektion auf den ersten Summanden und $\cdot kk(\pi)$ Rechtsmultiplikation mit $kk(\pi)$.

(b) Die Folge in (a) kann fortgesetzt werden zu einer exakten Folge

$$\begin{array}{ccccc} \xrightarrow{\cdot kk(\pi(0,1))} & kk_*(\mathfrak{D},\mathfrak{A}(0,1)) & \xrightarrow{\cdot kk(\alpha(0,1))} & kk_{*+1}(\mathfrak{D},\mathfrak{B}(0,1)) & \to \\ & kk_*(\mathfrak{D},C_\alpha) & \xrightarrow{\cdot kk(\pi)} & kk_*(\mathfrak{D},\mathfrak{A}) \xrightarrow{\cdot kk(\alpha)} & kk_*(\mathfrak{D},\mathfrak{B}) \end{array}$$

Beweis: (a) Das Element $z \in kk_*(\mathfrak{D},\mathfrak{A})$ sei durch den Homomorphismus $\varphi : J^{2n+*}\mathfrak{D} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ repräsentiert. Die Gleichung $\langle\varphi\rangle \cdot kk(\alpha) = 0$ bedeutet, dass für ein geeignetes $m \geq n$ die durch $\mathfrak{K}(\alpha)\circ\varphi$ induzierte Abbildung $J^{2m+*}\mathfrak{D} \to \mathfrak{K}\hat{\otimes}\mathfrak{B}$ über einen Homomorphismus $\gamma : J^{2m+*}\mathfrak{D} \to \mathfrak{K}\hat{\otimes}\mathfrak{B}[0,1)$ faktorisiert. Wir können annehmen,dass $m = n$. Das kommutative Diagramm

$$\begin{array}{ccc} J^{2n+*}\mathfrak{D} & \xrightarrow{\varphi} & \mathfrak{K}\hat{\otimes}\mathfrak{A} \\ \downarrow\gamma & & \downarrow\mathfrak{K}(\alpha) \\ \mathfrak{K}\hat{\otimes}\mathfrak{B}[0,1) & \longrightarrow & \mathfrak{K}\hat{\otimes}\mathfrak{B} \end{array}$$

definiert einen Homomorphismus $\gamma' : J^{2n+*}\mathfrak{D} \to \mathfrak{K}\hat{\otimes}C_\alpha$ so dass $\mathfrak{K}(\pi)\cdot\gamma' = \varphi$.
(b) Dies folgt wie üblich durch Iteration der Konstruktion in (a). Hierzu benutzt man die Tatsache, dass der Abbildungskegel C_π für die Projektion $\pi : C_\alpha \to \mathfrak{A}$ diffeotop zu $\mathfrak{B}(0,1)$ ist, und das folgende kommutative Diagramm

$$\begin{array}{ccc} C_\pi & \xrightarrow{\pi'} & C_\alpha \\ \uparrow & & \| \\ \mathfrak{B}(0,1) & \xrightarrow{\iota} & C_\alpha \end{array}$$

In diesem Diagramm ist ι die Inklusion von $\mathfrak{B}(0,1)$ in die zweite Komponente von C_α und der erste senkrechte Pfeil ist die erwähnte Diffeotopieäquivalenz (sie bildet $f \in \mathfrak{B}(0,1)$ auf $(\iota f, 0) \in C_\pi \subset C_\alpha \oplus \mathfrak{A}[0,1))$ ab.

Gleicherweise ist der Abbildungskegel C_ι für $\iota : \mathfrak{B}(0,1) \to C_\alpha$ enthalten in $\mathfrak{A}(0,1) \oplus \mathfrak{B}([0,1)\times[0,1))$. Die Projektion $C_\iota \to \mathfrak{A}(0,1)$ ist ebenfalls eine Diffeotopieäquivalenz und macht das folgende Diagramm kommutativ

$$\begin{array}{ccc} C_\iota & \longrightarrow & \mathfrak{B}(0,1) \\ \downarrow & & \| \\ \mathfrak{A}(0,1) & \xrightarrow{\alpha(0,1)} & \mathfrak{B}(0,1) \end{array}$$

q.e.d.

LEMMA 5.2 *$\alpha : \mathfrak{A} \to \mathfrak{B}$ und $\mathfrak{D}$ seien wie in 5.1*

(a) Die Folge

$$kk_*(C_\alpha,\mathfrak{D}) \xleftarrow{kk(\pi)\cdot} kk_*(\mathfrak{A},\mathfrak{D}) \xleftarrow{kk(\alpha)\cdot} kk_*(\mathfrak{B},\mathfrak{D})$$

ist exakt.

(b) Die Folge in (a) kann zu einer langen exakten Folge der Form

$$\begin{array}{ccccccc} \xleftarrow{kk(\pi(0,1))\cdot} & kk_*(\mathfrak{A}(0,1),\mathfrak{D}) & \xleftarrow{kk(\alpha(0,1))\cdot} & kk_{*+1}(\mathfrak{B}(0,1),\mathfrak{D}) & \leftarrow & & \\ & kk_*(C_\alpha,\mathfrak{D}) & \xleftarrow{kk(\pi)\cdot} & kk_*(\mathfrak{A},\mathfrak{D}) & \xleftarrow{kk(\alpha)\cdot} & kk_*(\mathfrak{B},\mathfrak{D}) & \end{array}$$

fortgesetzt werden.

Beweis: (a) Der Einfachheit halber nehmen wir an, dass $* = 0$. Sei dann $\varphi : J^{2n}\mathfrak{A} \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{D}$ ein Homomorphismus mit der Eigenschaft, dass $kk(\pi)\cdot\langle\varphi\rangle = \langle 0\rangle$. Dies bedeutet, dass ein kommutatives Diagramm der Form

$$\begin{array}{ccc} J^{2n}C_\alpha & \xrightarrow{J^{2n}(\pi)} & J^{2n}\mathfrak{A} \\ \downarrow\gamma & & \downarrow\varphi \\ \mathfrak{K}\hat{\otimes}\mathfrak{D}[0,1) & \xrightarrow{\mathrm{ev}} & \mathfrak{K}\hat{\otimes}\mathfrak{D} \end{array}$$

existiert. Hierbei ist ev die Auswertungsabbildung in 0. Man beachte, dass $\varphi\circ J^{2n}(\pi)\circ\varepsilon = \varphi\circ\varepsilon\circ J^{2n+2}(\pi)$, so dass wir annehmen können, dass die Diffeotopie schon auf Niveau n realisiert ist. Da γ in diesem Diagramm den Kern von $J^{2n}(\pi)$ in den Kern von ev abbildet, d.h. also in $\mathfrak{K}\hat{\otimes}\mathfrak{D}(0,1)$, ergibt die Einschränkung von γ eine Abbildung $\gamma' : J^{2n}(\mathfrak{B}(0,1)) \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{D}(0,1)$.

Wir verwenden jetzt die natürlichen Abbildungen $J\mathfrak{B} \to \mathfrak{B}(0,1)$ und $J(\mathfrak{D}(0,1)) \to \mathfrak{K}\hat{\otimes}\mathfrak{D}$, siehe 3.8, um durch die Komposition

$$J^{2n+2}\mathfrak{B} \longrightarrow J^{2n+1}(\mathfrak{B}(0,1)) \xrightarrow{J(\gamma')} J(\mathfrak{K}\hat{\otimes}\mathfrak{D}(0,1)) \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{D}$$

eine Abbildung $\psi : J^{2n+2}\mathfrak{B} \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{D}$ zu konstruieren. Wir müssen zeigen, dass $\psi\circ J^{2n+2}(\alpha) \sim \varphi\circ\varepsilon$. Dies folgt aus dem folgenden kommutativen Diagramm

$$\begin{array}{ccccccc} 0\longrightarrow & \mathfrak{K}\hat{\otimes}\mathfrak{D}(0,1) & \longrightarrow & \mathfrak{K}\hat{\otimes}\mathfrak{D}[0,1) & \longrightarrow & \mathfrak{K}\hat{\otimes}\mathfrak{D} & \longrightarrow 0 \\ & \uparrow\gamma' & & \uparrow\gamma & & \uparrow\varphi & \\ 0\longrightarrow & J^{2n}(\mathfrak{B}(0,1) & \longrightarrow & J^{2n}C_\alpha & \longrightarrow & J^{2n}\mathfrak{A} & \longrightarrow 0 \\ & \uparrow J^{2n}\alpha(0,1) & & \uparrow J^{2n}\alpha' & & \| & \\ 0\longrightarrow & J^{2n}(\mathfrak{A}(0,1)) & \longrightarrow & J^{2n}(\mathfrak{A}[0,1)) & \longrightarrow & J^{2n}\mathfrak{A} & \longrightarrow 0 \end{array}$$

Hierbei ist $\alpha(0,1)$ die Einhängung von α und α' ist die Abbildung, die $f\in\mathfrak{A}[0,1)$ auf $(f(0),\alpha[0,1)(f))\in C_\alpha$ abbildet.

Das Diagramm zeigt unter Verwendung von Lemma 3.3, dass die durch φ induzierte Abbildung $J^{2n+1}\mathfrak{A} \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{D}(0,1)$ diffeotop zur Komposition der folgenden Abbildungen ist

$$J^{2n+1}\mathfrak{A} \longrightarrow J^{2n}(\mathfrak{A}(0,1)) \xrightarrow{J^{2n}(\alpha(0,1))} J^{2n}(\mathfrak{B}(0,1)) \xrightarrow{\gamma'} \mathfrak{K}\hat{\otimes}\mathfrak{D}(0,1)$$

(b) folgt aus (a) genau wie in Lemma 5.1. *q.e.d.*

SATZ 5.3 *Es sei* $0\to\mathfrak{J}\to\mathfrak{A}\xrightarrow{q}\mathfrak{B}\to 0$ *eine linear zerfallende Erweiterung und* $e:\mathfrak{J}\to C_q$ *die Inklusionsabbildung, die durch* $e: x\mapsto(x,0)\in C_q\subset\mathfrak{A}\oplus\mathfrak{B}[0,1)$ *definiert ist. Dann ist* $kk(e)$ *ein invertierbares Element in* $kk_0(\mathfrak{J},C_q)$.

Beweis: Wir zeigen, dass das Inverse zu $kk(e)$ in $kk_0(C_q,\mathfrak{J})$ durch die Diffeotopieklasse $\langle u\rangle$ des Homomorphismus $u: J^2C_q\longrightarrow\mathfrak{K}\hat{\otimes}\mathfrak{J}$ gegeben ist, der folgendermaßen konstruiert wird: Sei $u_0: JC_q\longrightarrow\mathfrak{J}(0,1)$ die Abbildung, die zu der Erweiterung

$$0\longrightarrow\mathfrak{J}(0,1)\longrightarrow\mathfrak{A}[0,1)\longrightarrow C_q\longrightarrow 0$$

gehört. Dann sei u die Komposition von $J(u_0)$ mit der kanonischen Abbildung $J(\mathfrak{I}(0,1)) \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{I}$. Wir bezeichnen das durch u definierte Element auch mit $kk(u)$.

Das kommutative Diagramm

$$\begin{array}{ccccccccc}
0 \longrightarrow & \mathfrak{I}(0,1) & \longrightarrow & \mathfrak{A}[0,1) & \longrightarrow & C_q & \longrightarrow 0 \\
& \uparrow & & \uparrow & & \uparrow e & \\
0 \longrightarrow & \mathfrak{I}(0,1) & \longrightarrow & \mathfrak{I}[0,1) & \longrightarrow & \mathfrak{I} & \longrightarrow 0
\end{array}$$

zeigt, dass $u_0 \circ J(e)$ gerade die kanonische Abbildung $J\mathfrak{I} \longrightarrow \mathfrak{I}(0,1)$ ist, so dass also das Element $kk(e) \cdot kk(u)$ durch die Abbildung $\varepsilon : J^2\mathfrak{I} \longrightarrow \mathfrak{K}\hat{\otimes}\mathfrak{I}$ dargestellt wird. Nach Definition entspricht aber ε dem Einselement in $kk_0(\mathfrak{I},\mathfrak{I})$.

Um das umgekehrte Produkt $kk(u) \cdot kk(e)$ zu bestimmen, betrachten wir das kommutative Diagramm

$$\begin{array}{ccccccc}
0 \longrightarrow & C_q(0,1) & \longrightarrow & C_q[0,1) & \longrightarrow & C_q & \longrightarrow 0 \\
& \uparrow e(0,1) & & \uparrow e' & & \| & \\
0 \longrightarrow & \mathfrak{I}(0,1) & \longrightarrow & \mathfrak{A}[0,1) & \longrightarrow & C_q & \longrightarrow 0
\end{array} \tag{8}$$

wo

$$e'(f)(z) = \begin{cases} q(f(s)) & \text{wenn} \quad z = se^{i\theta} \quad \theta > 0 \,\text{und}\, s > 0 \\ 0 & \text{wenn} \quad |z| \geq 1 \\ f(s) & \text{wenn} \quad z = s \end{cases}$$

Hierbei werden Elemente von $C_q[0,1)$ aufgefasst als "Funktionen" g von zwei Variablen $(x,y) \in [0,1]^2$ oder von einer komplexen Variablen $z = x + iy$ mit

$$g(x+iy) \in \begin{cases} \mathfrak{A} & y = 0 \\ \mathfrak{B} & y > 0 \end{cases}$$

Außerdem muss eine Funktion g in $C_q[0,1)$ die folgenden Bedingungen erfüllen:

$g(x+iy) = 0$, wenn $x = 1$ oder $y = 1$

für $y > 0$ ist $g(x+iy)$ eine stetige Funktion von x, y

$q(g(x)) = \lim\limits_{y \to 0} g(x+iy)$

Das kommutative Diagramm (8) zeigt, dass $e(0,1) \circ u_0$ diffeotop zu der kanonischen Abbildung $JC_q \to C_q(0,1)$ ist und damit, dass $kk(u) \cdot kk(e) = 1$. *q.e.d.*

Betrachte nun die nach links unendlichen exakten Folgen aus 5.1(b) und 5.2(b) für den Fall, wo α die Quotientenabbildung q in einer Erweiterung wie in 5.3 ist. Theorem 5.3 erlaubt es, in den exakten Folgen jeweils C_q durch $\mathfrak{I}$ zu ersetzen. Überdies erhalten wir aus 5.3 auch sofort die Bottperiodizität.

SATZ 5.4 *Die durch die Einhängungserweiterung induzierten Abbildungen $J\mathfrak{A} \to \mathfrak{A}(0,1)$ und $J^2\mathfrak{A} \to \mathfrak{A}(0,1)^2$ repräsentieren in $kk_0(J\mathfrak{A},\mathfrak{A}(0,1))$, in $kk_1(\mathfrak{A},\mathfrak{A}(0,1))$ und in $kk_0(\mathfrak{A},\mathfrak{A}(0,1)^2)$ invertierbare Elemente.*

Beweis: Dies ergibt sich aus den langen exakten Folgen aus 5.1(b) und 5.2(b) angewandt auf das folgende kommutative Diagramm von Erweiterungen

$$\begin{array}{ccccccccc} 0 \longrightarrow & \mathfrak{A}(0,1) & \longrightarrow & \mathfrak{A}[0,1) & \longrightarrow & \mathfrak{A} & \longrightarrow 0 \\ & \uparrow & & \uparrow & & \uparrow & \\ 0 \longrightarrow & J\mathfrak{A} & \longrightarrow & T\mathfrak{A} & \longrightarrow & \mathfrak{A} & \longrightarrow 0 \end{array}$$

Z.B. zeigt das 5-Lemma und die exakte Folge aus 5.2(b) für $kk_0(\,\cdot\,, J\mathfrak{A})$, dass Linksmultiplikation mit $g = kk(J\mathfrak{A} \to \mathfrak{A}(0,1))$ einen Isomorphismus von $kk_0(\mathfrak{A}(0,1), J\mathfrak{A})$ mit $kk_0(J\mathfrak{A}, J\mathfrak{A})$ induziert. Man schließt daraus, dass g von rechts invertierbar ist. *q.e.d.*

THEOREM 5.5 *Es sei $\mathfrak{D}$ eine beliebige m-Algebra. Jede linear zerfallende Erweiterung*

$$E:\ 0 \to \mathfrak{J} \xrightarrow{i} \mathfrak{A} \xrightarrow{q} \mathfrak{B} \to 0$$

induziert exakte Folgen in $kk(\mathfrak{D}, \cdot\,)$ und $kk(\,\cdot\,, \mathfrak{D})$ der folgenden Form:

$$\begin{array}{ccccc} kk_0(\mathfrak{D}, \mathfrak{J}) & \xrightarrow{\cdot kk(i)} & kk_0(\mathfrak{D}, \mathfrak{A}) & \xrightarrow{\cdot kk(q)} & kk_0(\mathfrak{D}, \mathfrak{B}) \\ \uparrow & & & & \downarrow \\ kk_1(\mathfrak{D}, \mathfrak{B}) & \xleftarrow{\cdot kk(q)} & kk_1(\mathfrak{D}, \mathfrak{A}) & \xleftarrow{\cdot kk(i)} & kk_1(\mathfrak{D}, \mathfrak{J}) \end{array} \tag{9}$$

und

$$\begin{array}{ccccc} kk_0(\mathfrak{J}, \mathfrak{D}) & \xleftarrow{kk(q)\cdot} & kk_0(\mathfrak{A}, \mathfrak{D}) & \xleftarrow{kk(i)\cdot} & kk_0(\mathfrak{B}, \mathfrak{D}) \\ \downarrow & & & & \uparrow \\ kk_1(\mathfrak{B}, \mathfrak{D}) & \xrightarrow{kk(i)\cdot} & kk_1(\mathfrak{A}, \mathfrak{D}) & \xrightarrow{kk(q)\cdot} & kk_1(\mathfrak{J}, \mathfrak{D}) \end{array} \tag{10}$$

Die gegebene Erweiterung E definiert eine klassifizierende Abbildung $J\mathfrak{B} \to \mathfrak{J}$ und damit ein Element von $kk_1(\mathfrak{J}, \mathfrak{B})$, das wir mit $kk(E)$ bezeichnen. Die senkrechten Pfeile in (9) und (10) sind bis auf ein Vorzeichen gegeben durch Rechts-, bzw. durch Linksmultiplikation mit dieser Klasse $kk(E)$. Das Vorzeichen hängt von den Identifizierungen bei der Bottperiodizität nach Satz 5.4 ab.

Beweis: Satz 5.3 erlaubt es, in den exakten Folgen aus 5.1(b) und 5.2(b) jeweils C_q durch $\mathfrak{J}$ zu ersetzen. Dies ergibt unter Verwendung von 5.4 die exakten Folgen (9) und (10). Die Verbindungsabbildungen für die einfachen Einhängungen in 5.1(b) und 5.2(b) sind induziert durch die Inklusion $j : B(0,1) \to C_q$, d.h. sie sind gegeben durch Produkt mit der Klasse $kk(j)$. Das kommutative Diagramm

$$\begin{array}{ccccccc} 0 \to & \mathfrak{J}(0,1) & \to & \mathfrak{A}[0,1) & \to & C_q & \to 0 \\ & \| & & \uparrow & & \uparrow j & \\ 0 \to & \mathfrak{J}(0,1) & \to & \mathfrak{A}(0,1) & \to & B(0,1) & \to 0 \end{array}$$

zeigt andererseits, dass mit den Bezeichnungen aus dem Beweis zu Satz 5.3 die Identität $kk(j) \cdot kk(u) = kk(E)$ gilt. Die Identifikation von C_q mit $\mathfrak{J}$ geschieht aber gerade mit Hilfe des Isomorphismus, der nach Satz 5.3 durch Multiplikation mit $kk(e)^{-1} = kk(u) \in kk(C_q, \mathfrak{J})$ definiert ist. *q.e.d.*

BEMERKUNG 5.6 *Der Beweis für die Ausschneidung macht deutlich, dass in der Definition von kk die Beschränkung auf Diffeotopie, d.h. differenzierbare Homotopie als Äquivalenzrelation von grundlegender Bedeutung ist. Der Beweis von 5.2 und vor allem aber auch der zu 5.3 beruht auf der Existenz der Abbildung* $J(\mathfrak{A}(0,1)) \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$. *Wenn der Abbildungskegel* C_q *mit stetigen oder nur k-fach differenzierbaren Funktionen definiert worden wäre, würde das Inverse* $kk(u)$ *zu* $kk(e)$ *nicht existieren.*
Dies liegt an den Eigenschaften der Toeplitzerweiterung, bei der die Grösse des Ideals der des Quotienten entspricht. Man könnte verschiedene Versionen von kk definieren, indem man Diffeotopie durch stetige oder k-fach differenzierbare Homotopie ersetzt und dann aber auch statt der glatten Toeplizerweiterung entsprechend die stetige oder die k-fach differenzierbare Toeplitzerweiterung verwendet. Dies bedeutet, dass man in der Definition von kk das Ideal $\mathfrak{K}$ *durch die* C^**-Algebra* $\mathcal{K}$ *der kompakten Operatoren bzw. durch die Algebra* $\mathcal{K}_n$ *der Matrizen* (λ_{ij}) *mit*

$$\sum_{ij} |\lambda_{ij}|\,|1+i+j|^n \quad \leq \infty$$

ersetzen muss.

BEMERKUNG 5.7 *In Analogie zu [Sk] könnte man für zwei m-Algebren* $\mathfrak{A}$ *und* $\mathfrak{B}$ *eine Theorie* $kk_*^{nuk}(\mathfrak{A}, \mathfrak{B})$ *definieren, indem man statt beliebiger Homomorphismen nur nukleare Homomorphismen* $J^{2n+*}\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{B}$ *betrachtet. Aufgrund der Hochhebungs- und Fortsetzungseigenschaften nuklearer Abbildungen würde diese Theorie Ausschneidung in beiden Variablen für Erweiterungen von Fréchetalgebren erfüllen, auch wenn diese nicht notwendigerweise zerfallen.*

6 DER CHERN-CONNES-CHARAKTER

Wir zeigen in diesem Abschnitt, dass Funktoren E auf der Kategorie der m-Algebren, die gewisse abstrakte Eigenschaften besitzen, automatisch auch funktoriell unter kk-Elementen sind. Da die Definition von kk wesentlich auf der Periodizitätsabbildung ε beruht, besteht der erste Schritt darin, zu zeigen, dass für solche Funktoren $E(\varepsilon)$ ein Isomorphismus sein muss. Weil aber ε mit Hilfe der Toeplitzerweiterung definiert ist, benötigen wir zuerst eine genauere Analyse der universellen Eigenschaften der Toeplitzalgebra $\mathfrak{T}$. Hierzu sei $U(v,w)$ die universelle Algebra über $\mathbb{C}$ mit zwei Erzeugern v und w, die die Relation $wv = 1$ erfüllen. Dies ist Kurzschreibweise für die Bedingung, dass wv ein Einselement für alle Polynome in v und w ist. Wir setzen $e = 1-vw$. Dann ist e ein idempotentes Element in $U(v,w)$, und die Elemente $e^{ij} = v^i e w^j$ erfüllen

$$e^{ij}e^{kl} = \delta_{jk}e^{il}$$

Man sieht daraus sofort, dass man $U(v,w)$ treu auf dem Hilbertraum $\ell^2(\mathbb{N})$ mit der kanonischen Orthonormalbasis $(\xi_n)_{n=0,1,2,\ldots}$ durch

$$v\xi_n = \xi_{n+1} \qquad w\xi_n = \xi_{n-1},\ w\xi_0 = 0$$

darstellen kann. Dabei werden dann also die e^{ij} auf die Matrixeinheiten mit $e^{ij}\xi_n = \delta_{jn}\xi_i$ abgebildet. Die Linearkombinationen der $e^{ij}, 1 \leq i,j \leq n$ bilden eine Matrixalgebra isomorph zu $M_n(\mathbb{C})$

SATZ 6.1 *Die Toeplitzalgebra $\mathfrak{T}$ ist die universelle m-Algebra, die von zwei Elementen v und w mit $wv = 1$, erzeugt wird (d.h. also die eine Vervollständigung von $U(v,w)$ ist) und deren Topologie durch eine Familie $(p_n)_{n\in\mathbb{N}}$ von submultiplikativen Halbnormen bestimmt ist, die die folgende Wachstumsbedingung erfüllen*

$$p_n(v^k) \le C_n(1+k^n), \qquad k = 1,2,\ldots$$
$$p_n(w^k) \le C_n(1+k^n), \qquad k = 1,2,\ldots$$

Hierbei sind die C_n positive Konstanten.
Dies bedeutet, dass für jede m-Algebra $\mathfrak{B}$, deren Topologie durch einen Familie von Halbnormen $(p'_n)_{n\in\mathbb{N}}$ gegeben ist und die von zwei Elementen v' und w' erzeugt wird, die dieselben Relationen und Wachstumsbedingungen erfüllen, ein stetiger Homomorphismus $\mathfrak{T} \to \mathfrak{B}$ existiert, der v auf v' und w auf w' abbildet.

Beweis: Nach Definition ist $\mathfrak{T}$ als lokalkonvexer Vektorraum isomorph zu

$$\mathfrak{K} \oplus \mathcal{C}^\infty(S^1)$$

Wenn z den Erzeuger von $\mathcal{C}^\infty(S^1)$ bezeichnet, so entspricht unter diesem linearen Isomorphismus v^n dem Element z^n und w^n dem Element z^{-n}. Die in 1.5 angegeben Halbnormen erfüllen also offensichtlich die Wachstumsbedingung.

Sei $\mathfrak{B}$ wie in der Behauptung und φ der Homomorphismus $U(v,w) \to \mathfrak{B}$, der v auf v' und w auf w' abbildet. Es genügt zu zeigen, dass φ auf $U(v,w)\cap\mathfrak{K}$ und auf $U(v,w)\cap\mathcal{C}^\infty(S^1)$ stetig ist. Da

$$\begin{aligned} p'_n(\varphi(e^{ij})) &= p'_n(v'^i w'^j - v'^{i+1}w'^{j+1}) \\ &\le 2C'_n(1+(i+1)^n)C'_n(1+(j+1)^n) \le C(1+i+j)^n \end{aligned}$$

mit einer neuen Konstante C, ist φ auf dem ersten Summanden stetig und die Stetigkeit auf dem zweiten ist klar. *q.e.d.*

LEMMA 6.2 *(vgl. [Cu1, 4.2]) Es existieren eindeutig bestimmte stetige Homomorphismen $\varphi, \varphi' : \mathfrak{T} \to \mathfrak{T}\hat{\otimes}\mathfrak{T}$, so dass*

$$\varphi(v) = v(1-e)\otimes 1 + e\otimes v \qquad \varphi(w) = (1-e)w\otimes 1 + e\otimes w$$
$$\varphi'(v) = v(1-e)\otimes 1 + e\otimes 1 \qquad \varphi'(w) = (1-e)w\otimes 1 + e\otimes 1$$

Diese beiden Homomorphismen sind diffeotop und zwar durch eine Diffeotopie ψ_t : $\mathfrak{T} \to \mathfrak{T}\hat{\otimes}\mathfrak{T}$, $t\in[0,\pi/2]$, für die $\psi_t(x) - \varphi(x) \in \mathfrak{K}\hat{\otimes}\mathfrak{T}$ für alle $t\in[0,\pi/2], x\in\mathfrak{T}$ gilt.

Beweis: Wir zeigen, dass φ und φ' beide diffeotop zu ψ sind, wo

$$\psi(v) = v\otimes 1 \qquad \psi(w) = w\otimes 1$$

Wir schreiben im folgenden Linearkombinationen von $e^{ij}\otimes x$, $0\le i,j\le n-1, x\in\mathfrak{T}$ als $n\times n$-Matrizen mit Matrixelementen in $\mathfrak{T}$. Weiter schreiben wir E_n für $1-v^nw^n = e^{00}+e^{11}+\ldots+e^{n-1,n-1}$. Mit diesen Bezeichnungen setzen wir für $t\in[0,\pi/2]$

$$u_t = (1-E_2) + \begin{pmatrix} e+\cos t\,(1-e) & \sin t\, v \\ -\sin t\, w & \cos t\, 1 \end{pmatrix}$$
$$u'_t = (1-E_2) + \begin{pmatrix} \cos t\, 1 & \sin t\, 1 \\ -\sin t\, 1 & \cos t\, 1 \end{pmatrix}$$

Sowohl u_t als auch u'_t sind offensichtlich invertierbar in $\mathfrak{T}\hat{\otimes}\mathfrak{T}$. Wir zeigen nun, dass für jedes t stetige Homomorphismen $\varphi_t, \varphi'_t : \mathfrak{T} \to \mathfrak{T}\hat{\otimes}\mathfrak{T}$ existieren, so dass

$$\begin{aligned} \varphi_t(v) &= u_t\,(v\otimes 1) \qquad & \varphi_t(w) &= (w\otimes 1)\,u_t^{-1} \\ \varphi'_t(v) &= u'_t\,(v\otimes 1) \qquad & \varphi'_t(w) &= (w\otimes 1)\,u_t'^{-1} \end{aligned}$$

Seien $p_n \oplus q_n$ die Halbnormen aus 1.5, die die Topologie auf $\mathfrak{T}$ bestimmen. Wir müssen nachweisen, dass die Halbnormen $(p_n \oplus q_n)\otimes(p_n\oplus q_n)$ die Wachstumsbedingung auf den Potenzen von $u_t\,(v\otimes 1), (w\otimes 1)\,u_t^{-1}, u'_t\,(v\otimes 1)$ und $(w\otimes 1)\,u_t'^{-1}$ erfüllen.

Es ist nun aber

$$\big(u_t\,(v\otimes 1)\big)^k = u_t^{(k)}\,(v^k\otimes 1)$$

mit $u_t^{(k)} = (1-E_k)+L$, wo L eine invertierbare $k\times k$-Matrix mit Werten in $\mathfrak{T}$ ist. Man sieht sofort, dass L Summe von k^2 Elementen der Form $e^{ij}\otimes(\lambda_1 W_1 + \lambda_2 W_2)$, $0 \le i,j \le k$ ist, mit $|\lambda_i| \le 1$, W_i Wörter in v, w der Länge $\le k+1$. Daher gilt

$$\begin{aligned} p_n\otimes(p_n\oplus q_n)(L) &\le 2k^2 C_n(1+2k^n)(k+1)^n \le C(1+k^{2n+2}) \\ (p_n\oplus q_n)\otimes(p_n\oplus q_n)\big(u_t^{(k)}\,(v^k\otimes 1)\big) &\le C(1+k^{3n+2}) \end{aligned}$$

mit einer neuen Konstante C. Die Wachstumsbedingungen für $(p_n\oplus q_n)\otimes(p_n\oplus q_n)$ auf den Potenzen von $(w\otimes 1)\,u_t^{-1}, u'_t\,(v\otimes 1)$ und $(w\otimes 1)\,u_t'^{-1}$ ergeben sich im ersten Fall genauso und in den zwei letzteren sogar einfacher.

Die Familie φ_t ergibt nun eine Diffeotopie zwischen φ und ψ und die Familie φ'_t ergibt eine Diffeotopie zwischen φ' und ψ. Wir erhalten ψ_t durch Zusammensetzen dieser beiden Diffeotopien. Die geforderte Zusatzbedingung $\psi_t(x) - \varphi(x) \in \mathfrak{K}\hat{\otimes}\mathfrak{T}$ ist offensichtlich erfüllt. *q.e.d.*

Wir betrachten im folgenden Funktoren E von der Kategorie der m-Algebren in die Kategorie der abelschen Gruppen, die die folgenden (wohlbekannten) Bedingungen erfüllen:

(E1) E ist diffeotopieinvariant, d.h. die Auswertungsabbildung in einem beliebigen Punkt $t\in[0,1]$ induziert einen Isomorphismus $E(ev_t) : E(\mathfrak{A}[0,1]) \to E(\mathfrak{A})$

(E2) E ist stabil, d.h. die kanonische Inklusion $\iota : \mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ induziert einen Isomorphismus $E(\iota)$.

(E3) E ist halbexakt, d.h. jede linear zerfallende Erweiterung $0 \to \mathfrak{I} \to \mathfrak{A} \to \mathfrak{B} \to 0$ induziert eine kurze exakte Folge $E(\mathfrak{I}) \to E(\mathfrak{A}) \to E(\mathfrak{B})$

Wir erinnern daran, dass nach einer Standardkonstruktion aus der algebraischen Topologie die kurze exakte Folge in (E3) mit Hilfe von Abbildungskegeln und unter Benutzung der Eigenschaft (E1) zu einer nach links unendlichen langen exakten Folge der Form

$$\begin{aligned} \ldots \to E(\mathfrak{B}(0,1)^2) \to E(\mathfrak{I}(0,1)) \to E(\mathfrak{A}(0,1)) \\ \to E(\mathfrak{B}(0,1)) \to E(\mathfrak{I}) \to E(\mathfrak{A}) \to E(\mathfrak{B}) \end{aligned} \tag{11}$$

fortgesetzt werden kann, vgl. etwa [Ka] oder [Cu3].

Für eine m-Algebra $\mathfrak{A}$ seien $Q\mathfrak{A}, q\mathfrak{A}$ und $\iota, \bar{\iota} : \mathfrak{A} \to Q\mathfrak{A}$ wie in 2.4 definiert. Wir bezeichnen mit $\delta : q\mathfrak{A} \to \mathfrak{A}$ die Restriktion der Abbildung $Q\mathfrak{A} \to \mathfrak{A}$, die $\iota(x)$ auf x und $\bar{\iota}(x)$ auf 0 abbildet. Das folgende Lemma ist wohlbekannt in der Kategorie der C*-Algebren, vgl. [Cu2, 3.1].

LEMMA 6.3 *Es sei E ein Funktor mit den Eigenschaften (E1), (E2), (E3).*

(a) Die kanonische Abbildung $id{*}0 \oplus 0{*}id{:}Q\mathfrak{A} \longrightarrow \mathfrak{A} \oplus \mathfrak{A}$, die $\iota(x)$ auf $(x,0)$ und $\bar{\iota}(x)$ auf $(0,x)$ abbildet, induziert einen Isomorphismus $E(Q\mathfrak{A}) \to E(\mathfrak{A}) \oplus E(\mathfrak{A})$.

(b) Die Abbildung $\delta : q\mathfrak{A} \to \mathfrak{A}$ induziert einen Isomorphismus $E(\delta)$.

Beweis: (a) Man zeigt genau wie im Fall von C*-Algebren ([Cu2, 3.1]) unter Verwendung der universellen Eigenschaft des freien Produkts, dass die Komposition der angegebenen Abbildung mit der Abbildung

$$\mathfrak{A} \oplus \mathfrak{A} \longrightarrow \begin{pmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{A} \end{pmatrix} \subset M_2(Q\mathfrak{A})$$

in beide Richtungen diffeotop zu den kanonischen Einbettungen von $\mathfrak{A} \oplus \mathfrak{A}$ und $Q\mathfrak{A}$ in die 2×2-Matrizen über diesen Algebren ist.
(b) Dies folgt aus folgendem kommutativen Diagramm

$$\begin{array}{ccccccc} 0 \longrightarrow & q\mathfrak{A} & \longrightarrow & Q\mathfrak{A} & \xrightarrow{\pi} & \mathfrak{A} & \longrightarrow 0 \\ & \downarrow \delta & & \downarrow \varphi & & \downarrow \mathrm{id} & \\ 0 \longrightarrow & \mathfrak{A} & \longrightarrow & \mathfrak{A} \oplus \mathfrak{A} & \longrightarrow & \mathfrak{A} & \longrightarrow 0 \end{array}$$

(wo $\varphi = \mathrm{id} * 0 \oplus \pi$) in Kombination mit (a). *q.e.d.*

SATZ 6.4 *Sei $\mathfrak{T}_0$ der Kern der kanonischen stetigen Abbildung $\kappa : \mathfrak{T} \to \mathbb{C}$, die v und w auf 1 abbildet. Für jeden Funktor E mit den Eigenschaften (E1), (E2), (E3) und für jede m-Algebra $\mathfrak{A}$ gilt*

$$E(\mathfrak{T}_0 \hat{\otimes} \mathfrak{A}) = 0$$

Beweis: Wir betrachten erst den Fall $\mathfrak{A} = \mathbb{C}$ und benutzen hierzu die Homomorphismen $\mathfrak{T} \to \mathfrak{T} \hat{\otimes} \mathfrak{T}$ aus Lemma 6.2 und außerdem den Homomorphismus $\omega : \mathfrak{T} \to \mathfrak{T} \hat{\otimes} \mathfrak{T}$, der v auf $v(1-e) \otimes 1$ und w auf $(1-e)w \otimes 1$ abbildet. Die Homomorphismen $\psi_t * \omega : Q\mathfrak{T} \to \mathfrak{T} \hat{\otimes} \mathfrak{T}$ bilden $q\mathfrak{T}$ in $\mathfrak{K} \hat{\otimes} \mathfrak{T}$ ab und ergeben durch Restriktion eine Diffeotopie

$$\omega_t : q\mathfrak{T} \longrightarrow \mathfrak{K} \hat{\otimes} \mathfrak{T}$$

Nach Konstruktion von φ und φ' gilt

$$\omega_0 = \iota \circ \delta \qquad \omega_1 = \iota \circ j \circ \kappa \circ \delta$$

wobei κ wie oben, $j : \mathbb{C} \to \mathfrak{T}$ die kanonische Inklusion und $\delta : q\mathfrak{T} \to \mathfrak{T}$ die kanonische "Auswertung"abbildung ist. siehe 6.3. Nach 6.3 ist $E(\delta) : E(q\mathfrak{T}) \to E(\mathfrak{T})$ ein Isomorphismus. Da nach (E2) außerdem auch $E(\iota)$ ein Isomorphismus ist, folgt aus

$$E(\iota \circ \delta) = E(\iota \circ j \circ \kappa \circ \delta)$$

dass $E(j) \circ E(\kappa) = E(\mathrm{id}_{\mathfrak{T}})$. Da offensichtlich $\kappa \circ j = \mathrm{id}_{\mathbb{C}}$, sind also $E(j)$ und $E(\kappa)$ zueinander inverse Isomorphismen. Da die Erweiterung $0 \to \mathfrak{T}_0 \to \mathfrak{T} \to \mathbb{C} \to 0$ zerfällt, ergibt sich aus der langen exakten Folge (11) eine kurze exakte Folge

$$0 \longrightarrow E(\mathfrak{T}_0) \longrightarrow E(\mathfrak{T}) \xrightarrow{E(\kappa)} E(\mathbb{C}) \longrightarrow 0$$

wobei $E(\kappa)$ ein Isomorphismus ist. Dies zeigt, dass $E(\mathfrak{T}_0) = 0$. Der allgemeine Fall $E(\mathfrak{T}_0 \hat{\otimes} \mathfrak{A})$ ergibt sich durch Tensorieren aller Homomorphismen in dem eben gegebenen Beweis mit $\mathrm{id}_{\mathfrak{A}}$ oder durch Ersetzen von E durch $E(\,\cdot\,\hat{\otimes}\mathfrak{A})$. *q.e.d.*

Die Toeplitzerweiterung mit Ideal $\mathcal{K}_1$ und Quotienten $\mathcal{C}^1(S^1)$ wurde auch von Lafforgue untersucht. Für sie wurden in [La] Analoga zu Lemma 6.2 und Satz 6.4 bewiesen und daraus wie in [Cu1] gefolgert, dass jeder Funktor E' auf der Kategorie der Banachalgebren, der Eigenschaften analog zu (E1), E(2), E(3) hat, Bottperiodizität erfüllt. Das folgende Korollar ist ebenfalls eine Form der Bottperiodizität.

KOROLLAR 6.5 *Für jeden Funktor E auf der Kategorie der m-Algebren mit den Eigenschaften (E1), (E2), (E3) und für jede m-Algebra $\mathfrak{A}$ sind die Abbildungen $E(\varepsilon) : E(J^2\mathfrak{A}) \to E(\mathfrak{K}\hat{\otimes}\mathfrak{A})$ und $E(\varepsilon^n) : E(J^{2n}\mathfrak{A}) \to E(\mathfrak{K}\hat{\otimes}\mathfrak{A})$ Isomorphismen.*

Beweis: Betrachte die folgenden kommutativen Diagramme

$$\begin{array}{ccccccccc}
0 \longrightarrow & J^{2k+2}\mathfrak{A} & \longrightarrow & TJ^{2k+1}\mathfrak{A} & \longrightarrow & J^{2k+1}\mathfrak{A} & \longrightarrow 0 \\
& \downarrow \varepsilon & & \downarrow & & \downarrow & \\
0 \longrightarrow & \mathfrak{K}\hat{\otimes}\mathfrak{K}\hat{\otimes}\mathfrak{A} & \longrightarrow & \mathfrak{K}\hat{\otimes}\mathfrak{T}_0\hat{\otimes}\mathfrak{A} & \longrightarrow & \mathfrak{K}\hat{\otimes}\mathcal{C}^\infty(S^1 \setminus \{1\})\hat{\otimes}\mathfrak{A} & \longrightarrow 0
\end{array}$$

$$\begin{array}{ccccccccc}
0 \longrightarrow & J^{2k+1}\mathfrak{A} & \longrightarrow & TJ^{2k}\mathfrak{A} & \longrightarrow & J^{2k}\mathfrak{A} & \longrightarrow 0 \\
& \downarrow & & \downarrow & & \downarrow \varepsilon^k & \\
0 \longrightarrow & \mathfrak{K}\hat{\otimes}\mathfrak{A}(0,1) & \longrightarrow & \mathfrak{K}\hat{\otimes}\mathfrak{A}[0,1) & \longrightarrow & \mathfrak{K}\hat{\otimes}\mathfrak{A} & \longrightarrow 0
\end{array}$$

und die nach (11) mit diesen Erweiterungen assoziierten langen exakten Folgen. Die Gruppen $E(T\mathfrak{A})$, $E(\mathfrak{T}_0\hat{\otimes}\mathfrak{A})$ und $E(\mathfrak{A}[0,1))$ sind trivial für jede m-Algebra $\mathfrak{A}$, siehe 6.4. Außerdem ist die Inklusion $\mathfrak{K}\hat{\otimes}\mathfrak{A}(0,1) \to \mathfrak{K}\hat{\otimes}\mathfrak{A}\hat{\otimes}\mathcal{C}^\infty(S^1 \setminus \{1\})$ eine Diffeotopieäquivalenz (vgl. 1.1) und die Abbildung $E(\mathfrak{A}(0,1)) \to E(\mathfrak{A}\hat{\otimes}\mathcal{C}^\infty(S^1 \setminus \{1\}))$ ein Isomorphismus. Anwendung des 5-Lemmas zeigt dann, dass die senkrechten Pfeile auf der linken Seite unter E jeweils einen Isomorphismus induzieren, wenn dies für die Pfeile rechts der Fall ist. Die Behauptung ergibt sich dann durch Induktion nach k (mit $J^0\mathfrak{A} = \mathfrak{A}$ und $\varepsilon^0 = \iota$). *q.e.d.*

THEOREM 6.6 *Sei E ein kovarianter Funktor mit den Eigenschaften (E1), (E2), (E3). Dann kann mit jedem $h \in kk_0(\mathfrak{A}, \mathfrak{B})$ in eindeutiger Weise ein Morphismus $E(h) : E(\mathfrak{A}) \to E(\mathfrak{B})$ assoziert werden, so dass $E(h_1 \cdot h_2) = E(h_2) \circ E(h_1)$ und $E(kk(\alpha)) = E(\alpha)$ für jeden Homomorphismus $\alpha : \mathfrak{A} \to \mathfrak{B}$ zwischen m-Algebren. Die analoge Aussage gilt auch für kontravariante Funktoren.*

Beweis: Sei h durch $\eta : J^{2n}\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{B}$ repräsentiert. Wir setzen

$$E(h) = E(\iota)^{-1}E(\eta)E(\varepsilon^n)^{-1}$$

Zunächst einmal ist klar, dass $E(h)$ wohldefiniert ist und dass $E(kk(\alpha)) = E(\alpha)$. Die Verträglichkeit mit dem Produkt ergibt sich aus derselben Rechnung wie die im Beweis von Theorem 4.3 und ist eine Konsequenz von Lemma 4.2.
Die Eindeutigkeit schließlich ist offensichtlich. *q.e.d.*

Das vorhergehende Resultat erlaubt, wie im Fall der KK-Theorie für C*-Algebren, [Hi], [Bl] ein andere Interpretation. Hierzu bemerken wir, dass kk_0 als Kategorie aufgefasst werden kann, deren Objekte gerade die m-Algebren sind, und deren Morphismen zwischen $\mathfrak{A}$ und $\mathfrak{B}$ durch $kk_0(\mathfrak{A}, \mathfrak{B})$ gegeben sind. Diese Kategorie ist additiv in dem Sinn, dass die Morphismen zwischen zwei Objekten jeweils eine abelsche Gruppe bilden und dass das Produkt von Morphismen bilinear ist.
Wir bezeichnen den natürlichen Funktor von der Kategorie der m-Algebren in die Kategorie kk_0, der auf den Objekten die Identität ist, auch mit kk_0.

KOROLLAR 6.7 *Es sei F ein Funktor von der Kategorie der m-Algebren in eine additive Kategorie, deren Objekte ebenfalls die m-Algebren sind, mit $F(\beta \circ \alpha) = F(\alpha) \cdot F(\beta)$. Wir bezeichnen diese Kategorie ebenfalls mit F und ihre Morphismen mit $F(\mathfrak{A}, \mathfrak{B})$.*
Wir nehmen an, dass $F(\mathfrak{A}, \mathfrak{B})$ in der ersten Variablen als kontravarianter Funktor und in der zweiten Variablen als kovarianter Funktor jeweils die Eigenschaften (E1), (E2), (E3) erfüllt. Dann existiert ein eindeutig bestimmter kovarianter Funktor F' von der Kategorie kk_0 in die Kategorie F, so dass $F = F' \circ kk_0$.

Beweis: Wir zeigen zuerst, dass $\varepsilon^n : J^{2n}\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ für jede m-Algebra $\mathfrak{A}$ einen invertierbaren Morphismus $F(\varepsilon^n)$ induziert.
Da Links- und Rechtsmultiplikation mit $F(\cdot)$ für festgehaltene zweite oder erste Variable Funktoren in die Kategorie der abelschen Gruppen mit den Eigenschaften (E1), (E2), (E3) sind, existieren nach 4.5 und 6.6 Elemente x und y in $F(\mathfrak{A}, J^{2n}\mathfrak{A})$, so dass $x \cdot F(\varepsilon^n) = F(\mathrm{id}_{\mathfrak{A}})$ und $F(\varepsilon^n) \cdot y = F(\mathrm{id}_{J^{2n}\mathfrak{A}})$. Da dann x und y Links- und Rechtsinverse für $F(\varepsilon^n)$ sind, sind sie gleich und invers zu $F(\varepsilon^n)$.

Ebenso sieht man, dass $\iota : \mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ für jede Wahl von $\mathfrak{A}$ einen invertierbaren Morphismus $E(\iota)$ induziert. Wenn jetzt $h \in kk_0(\mathfrak{A}, \mathfrak{B})$ durch $\eta : J^{2n}\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{B}$ repräsentiert ist, können wir setzen

$$F'(h) = F(\iota)F(\varepsilon^n)^{-1}F(\eta)F(\iota)^{-1}$$

q.e.d.

Auf der Kategorie der m-Algebren ist also kk_0 der universelle Funktor in eine additive Kategorie mit den Eigenschaften (E1), (E2), (E3) in beiden Variablen. Hieraus ergibt sich als Spezialfall sofort die Existenz des bivarianten Chern-Connes-Charakters im geraden Fall. Wir fassen hierzu die bivariante periodische zyklische Theorie $HP^0(\cdot, \cdot)$ ebenso wie kk_0 als additive Kategorie, deren Objekte die m-Algebren sind, auf. Ebenso wie bei kk schreiben wir das Produkt in HP^* in der umgekehrten Reihenfoge wie bei Homomorphismen. Für einen Homomorphismus α bezeichnen wir mit $ch(\alpha)$ das entsprechende Element der bivarianten zyklischen Theorie.

KOROLLAR 6.8 *Es existiert ein eindeutig bestimmter (kovarianter) Funktor ch :* $kk_0 \to HP^0$*, so dass* $ch(kk(\alpha)) = ch(\alpha) \in HP^0(\mathfrak{A}, \mathfrak{B})$ *für jeden Homomorphismus* $\alpha : \mathfrak{A} \to \mathfrak{B}$ *zwischen m-Algebren.*

Beweis: Die Eigenschaften (E1) und (E2) sind für die beiden Variablen von HP^0 seit langem bekannt und im wesentlichen schon von Connes in [Co] bewiesen. Der Nachweis von Eigenschaft (E3) gelang in [CuQu2]. *q.e.d.*

Der Chern-Connes-Charakter ch ist also eine bilineare multiplikative Transformation von kk_0 nach HP^0. Offensichtlich respektiert er auch das äußere Produkt auf kk_0 aus 4.3 (b), bzw. auf HP^0, siehe [CuQu2, p.86]. Es bleibt noch die Aufgabe, ch zu einer multiplikativen Transformation von der $\mathbb{Z}/2$-graduierten Theorie kk_* nach HP^* auszudehnen und die Verträglichkeit von ch mit der Randabbildung in den langen exakten Folgen zu untersuchen.
Wenn $E:\ 0 \to \mathfrak{I} \to \mathfrak{A} \to \mathfrak{B} \to 0$ eine linear zerfallende Erweiterung ist, schreiben wir wie in [CuQu2] δ für die Randabbildung $HP^i(\mathfrak{I}, \mathfrak{D}) \to HP^{i-1}(\mathfrak{B}, \mathfrak{D})$ in der ersten Variable und δ' für die Randabbildung $HP^i(\mathfrak{D}, \mathfrak{B}) \to HP^{i+1}(\mathfrak{D}, \mathfrak{I})$ in der zweiten Variable. Weiter schreiben wir im folgenden $1_{\mathfrak{A}}$ für $ch(\mathrm{id}_{\mathfrak{A}}) \in HP^0(\mathfrak{A}, \mathfrak{A})$.

Man rechnet leicht nach, dass $\delta'(1_{\mathfrak{B}}) = -\delta(1_{\mathfrak{I}})$, siehe [CuQu2,5.4]. Wie in [CuQu2] bezeichnen wir dieses Element von $HP^1(\mathfrak{B}, \mathfrak{I})$ mit $ch(E)$.

Ein Teil des folgenden Satzes wurde in etwas anderer Weise schon in [Ni1], [Ni2] bewiesen. Der Faktor $2\pi i$ beim Vergleich der Periodizitätsabbildungen in der K-Theorie und der zyklischen Homologie wurde an verschiedenen Stellen in der Literatur bemerkt, [Co1], [Pu], [Ni1].

SATZ 6.9 *Wir betrachten die Einhängungserweiterung*

$$E_\sigma : \qquad 0 \to \mathbb{C}(0,1) \to \mathbb{C}(0,1] \to \mathbb{C} \to 0$$

und die Toeplitzerweiterung

$$E_\tau : \qquad 0 \to \mathfrak{K} \to \mathfrak{T} \to \mathcal{C}^\infty S^1 \to 0$$

sowie die Einbettungsabbildungen $j : \mathbb{C}(0,1) \to \mathcal{C}^\infty S^1$ *und* $\iota : \mathbb{C} \to \mathfrak{K}$*. Mit dem Produkt in* HP^* *gilt die folgende fundamentale Beziehung*

$$ch(E_\sigma) \cdot ch(j) \cdot ch(E_\tau) = \frac{1}{2\pi i} ch(\iota)$$

Beweis: Wir benutzen kanonische dichte Unteralgebren von $\mathbb{C}(0,1), \mathcal{C}^\infty S^1, \mathfrak{K}$ und $\mathfrak{T}$ sowie ihre algebraische periodische zyklische Homologie HP_*^{alg}. Außerdem benutzen wir, wenn B eine dieser Algebren ist, die Homologie $HX_*(B)$ des X-Komplexes

$$X(B): \qquad B \quad \underset{b}{\overset{d}{\rightleftarrows}} \quad \Omega^1(B)_\natural$$

Wir haben zwei große kommutative Diagramme, wo die horizontalen Abbildungen alle Isomorphismen sind und die Spalten exakte Folgen mit 6 Termen

$$\begin{array}{ccccc}
\downarrow & & \downarrow & & \downarrow \\
HP_*(\mathfrak{K}) & \overset{\cong}{\longleftarrow} & HP_*^{alg}(M_\infty) & \overset{\cong}{\longrightarrow} & HX_*(M_\infty) \\
\downarrow & & \downarrow & & \downarrow \\
HP_*(\mathfrak{T}) & \overset{\cong}{\longleftarrow} & HP_*^{alg}(U(v,w)) & \overset{\cong}{\longrightarrow} & HX_*(U(v,w)) \\
\downarrow & & \downarrow & & \downarrow \\
HP_*(\mathcal{C}^\infty(S^1)) & \overset{\cong}{\longleftarrow} & HP_*^{alg}(\mathbb{C}[z,z^{-1}]) & \overset{\cong}{\longrightarrow} & HX_*(\mathbb{C}[z,z^{-1}]) \\
\downarrow & & \downarrow & & \downarrow
\end{array}$$

Die Folge in der rechten Spalte ist exakt, weil M_∞ H-unital und damit $HX_*(M_\infty)$ isomorph zu der Homologie $HX_*(M_\infty : U(v,w))$ des relativen X-Komplexes ist, vgl. [Wo]. Der Isomorphismus $HP_*^{alg}(M_\infty) \cong HX_*(M_\infty)$ gilt, weil M_∞ quasifrei ist, siehe [CuQu1,5.4]. Die Abbildungen in der mittleren Zeile sind Isomorphismen nach dem 5-Lemma. Das zweite Diagramm ist das folgende

$$\begin{array}{ccccc}
\downarrow & & \downarrow & & \downarrow \\
HP_*(\mathcal{C}_0^\infty(0,1)) & \overset{\cong}{\longleftarrow} & HP_*^{alg}((t-t^2)\mathbb{C}[t]) & \overset{\cong}{\longrightarrow} & HX_*((t-t^2)\mathbb{C}[t] : t\mathbb{C}[t]) \\
\downarrow & & \downarrow & & \downarrow \\
HP_*(\mathcal{C}_0^\infty(0,1]) & \overset{\cong}{\longleftarrow} & HP_*^{alg}(t\mathbb{C}[t]) & \overset{\cong}{\longrightarrow} & HX_*(t\mathbb{C}[t]) \\
\downarrow & & \downarrow & & \downarrow \\
HP_*(\mathbb{C}) & \overset{\cong}{\longleftarrow} & HP_*^{alg}(\mathbb{C}) & \overset{\cong}{\longrightarrow} & HX_*(\mathbb{C}) \\
\downarrow & & \downarrow & & \downarrow
\end{array}$$

Hierbei bezeichnen $\mathcal{C}_0^\infty(0,1]$ und $\mathcal{C}_0^\infty(0,1)$ die Algebren der glatten Funktionen auf $[0,1]$, die bei 0, bzw. bei 0 und 1 verschwinden (ohne Bedingung an die Ableitungen) und $HX_*((t-t^2)\mathbb{C}[t] : t\mathbb{C}[t])$ bezeichnet wieder die Homologie des relativen X-Komplexes. Die Isomorphismen in der ersten Zeile gelten nach dem 5-Lemma.
Um die Randabbildungen in der Toeplitz- und Einhängungserweiterung in der Spalte ganz links zu bestimmen, genügt es daher, die Randabbildungen in der Spalte ganz rechts zu berechnen. Dies ist aber sehr einfach. Nach Definition genügt es, jeweils Urbilder für die Repräsentanten einer Klasse in dem Komplex in der Mitte zu finden und dann den Randoperator des X-Komplexes darauf anzuwenden. Dies ergibt Elemente des relativen Komplexes, die das Bild unter der Randabbildung darstellen.
Fangen wir mit dem Erzeuger von $HX_0(\mathbb{C})$ an. Er wird durch $1 \in \mathbb{C}$ repräsentiert. Ein Urbild in $X_0(t\mathbb{C}[t])$ ist t. Unter dem Randoperator d geht dies auf $\natural(dt) \in \Omega^1((t-t^2)\mathbb{C}[t] : t\mathbb{C}[t])_\natural$.
Die Klasse von $\natural(dt)$ wiederum entspricht unter den Identifizierungen

$$\begin{aligned}
HX_1((t-t^2)\mathbb{C}[t] : t\mathbb{C}[t]) \leftarrow HP_1^{alg}((t-t^2)\mathbb{C}[t]) \rightarrow HP_1(\mathcal{C}_0^\infty(0,1)^\sim) \\
\leftarrow HP_1(\mathcal{C}^\infty(S^1)) \leftarrow HX_1(\mathbb{C}[z,z^{-1}])
\end{aligned}$$

der Klasse von $\natural(\frac{1}{2\pi i}z^{-1}dz)$. In der Tat ist $z = e^{2\pi it}$ und in den Differentialformen über S^1 ist $z^{-1}dz = 2\pi i\,dt$ (man beachte, dass $HP_1(\mathcal{C}^\infty(S^1))$ durch die de Rham

Kohomologie von S^1 gegeben ist). Ein Urbild für $\natural(z^{-1}dz)$ in $X_1(U(v,w))$ ist $\natural(wdv)$. Unter der Randabbildung des X-Komplexes wird $\natural(wdv)$ auf $b(wdv) = wv - vw = e$ abgebildet. *q.e.d.*

Für den speziellen Fall der universellen Erweiterung

$$E_u : \qquad 0 \to J\mathfrak{A} \to T\mathfrak{A} \to \mathfrak{A} \to 0$$

setzen wir

$$x_{\mathfrak{A}} = ch(E_u) = \delta'(1_{\mathfrak{A}}) = -\delta(1_{J\mathfrak{A}}) \in HP^1(\mathfrak{A}, J\mathfrak{A})$$

Die Randabbildungen δ und δ' in der universellen Erweiterung sind durch Links- und Rechtsmultiplikation mit $x_{\mathfrak{A}}$ gegeben. Die Tatsache, dass δ und δ' für die universelle Erweiterung Isomorphismen sind, impliziert sofort, dass $x_{\mathfrak{A}}$ invertierbar ist (es existieren Elemente y und y' in $HP^1(J\mathfrak{A}, \mathfrak{A})$ so dass $\delta(y) = x_{\mathfrak{A}} \cdot y = 1_{\mathfrak{A}}$ und $\delta'(y') = y' \cdot x_{\mathfrak{A}} = 1_{J\mathfrak{A}}$). Falls δ und δ' wieder die Randabbildungen in den exakten Folgen zu einer beliebigen Erweiterung $E: \ 0 \to \mathfrak{I} \to \mathfrak{A} \to \mathfrak{B} \to 0$ sind und $\alpha : J\mathfrak{B} \to \mathfrak{I}$ die klassifizierende Abbildung, so gilt wegen der Natürlichkeit der Randabbildung, dass

$$\delta(1_{\mathfrak{I}}) = x_{\mathfrak{B}} \cdot ch(\alpha) \qquad\qquad \delta'(1_{\mathfrak{B}}) = ch(\alpha) \cdot x_{\mathfrak{B}} \tag{12}$$

d.h. also $ch(E) = x_{\mathfrak{B}} \cdot ch(\alpha)$, siehe auch [CuQu2, 5.5]. Weiter gilt für jeden Homomorphismus $\alpha : J\mathfrak{A} \to \mathfrak{B}$

$$ch(\alpha) \cdot x_{\mathfrak{B}} \ = \ x_{J\mathfrak{A}} \cdot ch(J(\alpha)) \tag{13}$$

SATZ 6.10 *Sei $\varepsilon : J^2\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ die kanonische Abbildung. Dann gilt*

$$x_{\mathfrak{A}} \cdot x_{J\mathfrak{A}} \cdot ch(\varepsilon) \cdot ch(\iota)^{-1} \ = \ \frac{1}{2\pi i}\, 1_{\mathfrak{A}}$$

Beweis: Wir betrachten zuerst den Fall $\mathfrak{A} = \mathbb{C}$. Die Abbildung ε kann geschrieben werden als $\varepsilon = \varepsilon_2 \circ J(j) \circ J(\varepsilon_1)$, wo $\varepsilon_1 : J(\mathbb{C}) \to \mathbb{C}(0,1)$ und $\varepsilon_2 : J(\mathcal{C}^\infty S^1) \to \mathfrak{K}$ die klassifizierenden Abbildungen für die Einhängungs- und für die Toeplitzerweiterung sind und $j : \mathbb{C}(0,1) \to \mathcal{C}^\infty S^1$ die Einbettungsabbildung bezeichnet. Daher

$$x_{\mathfrak{A}} \cdot x_{J\mathfrak{A}} \cdot ch(\varepsilon) = x_{\mathfrak{A}} \cdot ch(\varepsilon_1) \cdot ch(j) \cdot x_{J\mathfrak{A}} \cdot ch(\varepsilon_2) = ch(E_\sigma) \cdot ch(j) \cdot ch(E_\tau)$$

Die erste Gleichung gilt nach (13) und die zweite folgt aus (12). Die Behauptung für $\mathfrak{A} = \mathbb{C}$ reduziert sich daher auf Satz 6.9.

Für allgemeines $\mathfrak{A}$ gilt unter Verwendung des äußeren Produkts in HP^* (siehe [CuQu2, p.86])

$$\begin{aligned} &x_{\mathfrak{A}} \cdot x_{J\mathfrak{A}} \cdot ch(\varepsilon_{\mathfrak{A}}) \cdot ch(\iota_{\mathfrak{A}})^{-1} = (x_{\mathbb{C}} \otimes 1_{\mathfrak{A}}) \cdot (x_{J\mathbb{C}} \otimes 1_{\mathfrak{A}}) \cdot (ch(\varepsilon_{\mathbb{C}}) \otimes 1_{\mathfrak{A}}) \\ &= (x_{\mathbb{C}} \cdot x_{J\mathbb{C}} \cdot ch(\varepsilon_{\mathbb{C}}) \cdot ch(\iota_{\mathbb{C}})^{-1}) \otimes 1_{\mathfrak{A}} = \tfrac{1}{2\pi i}\, 1_{\mathfrak{A}} \end{aligned}$$

q.e.d.

Sei jetzt u ein Element in $kk_1(\mathfrak{A},\mathfrak{B})$. Nach Definition ist $kk_1(\mathfrak{A},\mathfrak{B}) = kk_0(J\mathfrak{A},\mathfrak{B})$. Sei u_0 das Element in $kk_0(J\mathfrak{A},\mathfrak{B})$, das u entspricht. Wir setzen

$$ch(u) = \sqrt{2\pi i}\, x_{\mathfrak{A}} \cdot ch(u_0) \qquad \in HP^1(\mathfrak{A},\mathfrak{B})$$

SATZ 6.11 *Der so definierte Chern-Connes-Charakter ist multiplikativ, d.h. für* $u \in kk_i(\mathfrak{A},\mathfrak{B})$ *und* $v \in kk_j(\mathfrak{B},\mathfrak{C})$ *gilt*

$$ch(u \cdot v) = ch(u) \cdot ch(v)$$

Beweis: Der einzig wirklich neue Fall ist $i = j = 1$. Wir haben nach Lemma 6.9

$$ch(u) \cdot ch(v) = 2\pi i\, x_{\mathfrak{A}} \cdot ch(u_0) \cdot x_{\mathfrak{B}} \cdot ch(v_0) = 2\pi i\, x_{\mathfrak{A}} \cdot x_{J\mathfrak{A}} \cdot ch(J(u_0)) \cdot ch(v_0)$$

und andererseits nach Definition von ch im geraden Fall

$$ch(u \cdot v) = ch(\iota) \cdot ch(\varepsilon)^{-1} \cdot ch(J(u_0) \cdot v_0)$$

Die beiden Ausdrücke stimmen nach Satz 6.10 überein. *q.e.d.*

Insbesondere ist der Chern-Connes-Charakter auch mit den Randabbildungen in den langen exakten Folgen in kk_* und HP^*, die mit einer linear zerfallenden Erweiterung

$$(E) \qquad 0 \to \mathfrak{I} \to \mathfrak{A} \to \mathfrak{B} \to 0$$

von m-Algebren assoziiert sind, (bis auf den Faktor $\sqrt{2\pi i}$ und möglicherweise ein Vorzeichen) verträglich: Die klassifizierende Abbildung $J\mathfrak{B} \to \mathfrak{I}$ ergibt Elemente $kk(E) \in kk_1(\mathfrak{B},\mathfrak{I})$ und $ch(E) \in HP^1(\mathfrak{B},\mathfrak{I})$. Nach Definition gilt $\sqrt{2\pi i} ch(E) = ch(kk(E))$. Die Randabbildungen in den langen exakten Folgen in kk und HP sind laut Theorem 5.5 und [CuQu, 5.5] bis auf ein Vorzeichen durch Multiplikation mit $kk_1(E)$ bzw. $ch(E)$ gegeben.

Wir diskutieren jetzt zum Schluss noch den Zusammenhang mit dem Chern-Connes-Charakter, der für p-summierbare Fredholm- und Kasparovmoduln von Connes, Nistor und anderen konstruiert wurde, [Co], [Ni1].

SATZ 6.12 *Gegeben seien m-Algebren $\mathfrak{I}$ und $\mathfrak{A}$. Wir nehmen an, dass stetige Abbildungen $\alpha : \mathfrak{I} \to \mathfrak{A}$ und $\mu : \mathfrak{A}\hat{\otimes}\mathfrak{A} \to \mathfrak{I}$ mit folgenden Eigenschaften existieren:*

(a) $\alpha \circ \mu$ ist die Multiplikation auf $\mathfrak{A}$

(b) $\mu \circ (\alpha \otimes \alpha)$ ist die Multiplikation auf $\mathfrak{I}$

(insbesondere ist also $\alpha(\mathfrak{I})$ ein Ideal in $\mathfrak{A}$ mit $\mathfrak{A}^2 \subset \alpha(\mathfrak{I})$). Dann ist $kk(\alpha)$ ein invertierbares Element in $kk_0(\mathfrak{I},\mathfrak{A})$.

Beweis: Das Inverse zu $kk(\alpha)$ ist durch die Zusammensetzung der Toeplitzerweiterung mit der folgenden Erweiterung bestimmt

$$0 \to \mathfrak{I}(0,1) \to \mathfrak{I}(0,1) + \mathfrak{A}t \to \mathfrak{A} \to 0 \tag{14}$$

Die m-Algebra $\mathfrak{I}(0,1)+\mathfrak{A}t$ ist folgendermaßen definiert. Als lokalkonvexer Vektorraum ist sie einfach die direkte Summe von $\mathfrak{I}(0,1)$ und $\mathfrak{A}$. Das Symbol t bezeichnet die identische Funktion auf $[0,1]$. Die Elemente von $\mathfrak{A}t$ werden als Funktionen auf $[0,1]$ mit Werten in $\mathfrak{A}$, die Vielfache dieser Funktion mit Elementen von $\mathfrak{A}$ sind, aufgefasst. Die Multiplikation auf dem ersten Summanden ist die von $\mathfrak{J}(0,1)$. Das Produkt einer Funktion f in $\mathfrak{J}(0,1)$ mit einem Element $xt \in \mathfrak{A}t$ ist $\mu(\alpha(f) \otimes xt)$ (wir setzen hier μ und α kanonisch auf Funktionen fort). Das Produkt von xt und yt in dem zweiten Faktor ist definiert als $\mu(x \otimes y)(t^2 - t) + \alpha\mu(x \otimes y)t$, wobei der erste Summand in $\mathfrak{J}(0,1)$ und der zweite in $\mathfrak{A}t$ liegt. Man prüft sofort nach, dass mit diesen Definitionen $\mathfrak{J}(0,1) + \mathfrak{A}t$ eine m-Algebra ist.

Die Erweiterung (14) ist dann offensichtlich linear zerfallend und definiert ein Element u in $kk_1(\mathfrak{A}, \mathfrak{J}(0,1))$. Wir müssen nachweisen, dass das Produkt von u mit α in beide Richtungen die kanonischen Abbildungen $J\mathfrak{A} \to \mathfrak{A}(0,1)$ und $J\mathfrak{J} \to \mathfrak{J}(0,1)$ ergibt. Betrachte hierzu das folgende kommutative Diagramm

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \mathfrak{J}(0,1) & \longrightarrow & \mathfrak{J}(0,1] & \longrightarrow & \mathfrak{J} & \longrightarrow & 0 \\
 & & \downarrow \mathrm{id} & & \downarrow \mathrm{id}+\alpha & & \downarrow \alpha & & \\
0 & \longrightarrow & \mathfrak{J}(0,1) & \longrightarrow & \mathfrak{J}(0,1) + \mathfrak{A}t & \longrightarrow & \mathfrak{A} & \longrightarrow & 0 \\
 & & \downarrow \alpha & & \downarrow \alpha & & \| & & \\
0 & \longrightarrow & \mathfrak{A}(0,1) & \longrightarrow & \mathfrak{A}(0,1] & \longrightarrow & \mathfrak{A} & \longrightarrow & 0
\end{array}$$

Man beachte, dass die in der offensichtlichen Weise definierte Abbildung $\mathrm{id} + \alpha$ nach Bedingung (b) ein Homomorphismus ist. Der obere Teil des Diagramms zeigt nach Lemma 3.3, dass das Produkt $kk(\alpha) \cdot u$ durch die Einhängungserweiterung von $\mathfrak{J}$ repräsentiert wird, während der untere Teil zeigt, dass $u \cdot kk(\alpha)$ die Einhängungserweiterung von $\mathfrak{A}$ ist. *q.e.d.*

Wir können dieses Resultat nun anwenden auf die Schattenideale $\ell^p = \ell^p(H)$. Betrachte allgemeiner den Fall, wo $\mathfrak{J} = \ell^p \hat{\otimes} \mathfrak{B}$ und $\mathfrak{A} = \ell^q \hat{\otimes} \mathfrak{B}$ für eine beliebige m-Algebra $\mathfrak{B}$ und $p \leq q \leq 2p$. Die Abbildungen α und μ ergeben sich durch die Inklusion $\ell^p \to \ell^q$ und die Multiplikationsabbildung $\ell^q \hat{\otimes} \ell^q \to \ell^p$.

Satz 6.11 zeigt, dass $\ell^p \hat{\otimes} \mathfrak{B}$ und $\ell^q \hat{\otimes} \mathfrak{B}$ äquivalent in kk_0 und damit auch in HP^0 sind. Durch Iteration ist $\ell^p \hat{\otimes} \mathfrak{B}$ äquivalent zu $\ell^1 \hat{\otimes} \mathfrak{B}$ für jedes $p \geq 1$. Andererseits ist $\ell^1 \hat{\otimes} \mathfrak{B}$ in HP_0 äquivalent zu $\mathfrak{B}$, siehe etwa [Ga]. Wir erhalten also

KOROLLAR 6.13 *Die m-Algebra $\ell^p \hat{\otimes} \mathfrak{B}$ ist in HP_0 äquivalent zu $\mathfrak{B}$ für jedes $p \geq 1$. Der Chern-Connes-Charakter gibt eine Transformation $ch^{(p)} : kk_*(\mathfrak{A}, \ell^p \hat{\otimes} \mathfrak{B}) \to HP^*(\mathfrak{A}, \mathfrak{B})$ mit der Eigenschaft, dass*

$$ch^{(p)}(x \cdot kk(\iota^{(p)})) = ch(x) \qquad \text{für } x \in kk_*(\mathfrak{A}, \mathfrak{B})$$

wo $\iota^{(p)}$ die kanonische Inklusion $\mathfrak{B} \to \ell^p \hat{\otimes} \mathfrak{B}$ bezeichnet.

Durch Vergleich der funktoriellen Eigenschaften [Ni1, Theorem 3.5] sieht man ohne weiteres, dass dieser Chern-Connes-Charakter mit dem von Connes und Nistor konstruierten Charakter übereinstimmen muß.

7 VERGLEICH MIT DER TOPOLOGISCHEN K-THEORIE

Wir untersuchen in diesem Abschnitt den Spezialfall des Funktors kk, wo die erste Variable trivial ist, d.h. also $kk_*(\mathbb{C}, \cdot)$. Wir zeigen, dass dieser Funktor mit der topologischen K-Theorie übereinstimmt - im wesentlichen, wann immer diese definiert ist. Dazu benutzen wir die von Phillips eingeführte Theorie [Ph], die die topologische K-Theorie für die bisher wohl größte Klasse von lokalkonvexen Algebren, nämlich für m-Algebren, die gleichzeitig Fréchetalgebren sind, definiert. Dies hat für uns den Vorteil, dass diese Theorie es erlaubt, den Funktor K_* direkt auch auf Algebren vom Typ $J^n\mathbb{C}$ usw., die die Grundlage unserer Theorie bilden, anzuwenden. Dies vereinfacht den Beweis für Theorem 7.4 (selbst im Fall $\mathfrak{A} = \mathbb{C}$) bedeutend. Wir skizzieren am Ende des Abschnitts kurz, wie Theorem 7.4 ohne Verwendung der Theorie von Phillips für spezielle Fréchetalgebren, nämlich abgeleitete Unteralgebren von Banachalgebren bewiesen werden kann. Damit erhält man einen neuen Zugang zur K-Theorie von m-Algebren, indem man einfach $K_*(\mathfrak{A}) = kk_*(\mathbb{C}, \mathfrak{A})$ setzt.

Wie Phillips verstehen wir in dieser Arbeit unter Fréchetalgebren immer Fréchetalgebren, die auch m-Algebren sind, d.h. also vollständige lokalkonvexe Algebren, deren Topologie durch eine abzählbare Familie von submultiplikativen Halbnormen bestimmt ist.

Mit einer Fréchetalgebra $\mathfrak{A}$ assoziiert Phillips in [Ph] die folgende abelsche Gruppe:

$$\begin{aligned} K_0(\mathfrak{A}) = \{\, [e] \,|\, & e \text{ ist ein idempotentes Element in} \\ & M_2(\mathfrak{K}\hat{\otimes}\mathfrak{A}^\sim) \text{ so dass } e - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \in M_2(\mathfrak{K}\hat{\otimes}\mathfrak{A})\} \end{aligned} \tag{15}$$

Hierbei bezeichnet, wie üblich $\mathfrak{K}\hat{\otimes}\mathfrak{A}^\sim$ die Algebra, die man erhält, wenn man zu $\mathfrak{K}\hat{\otimes}\mathfrak{A}$ eine Eins adjungiert.

Phillips verwendet die Bezeichnung "RK_0" für diese Gruppe. Uns erscheint die Bezeichnung K_0 angemessener, da diese Theorie die übliche topologische K-Theorie von der Kategorie der Banachalgebren auf die der Fréchetalgebren verallgemeinert. Wir setzen auch $K_1(\mathfrak{A}) = K_0(\mathfrak{A}(0,1))$

In (15) bezeichnet $[e]$ die Homotopieklasse von e. In [Ph] wird gezeigt, dass zwei idempotente Elemente e und e' in $M_2(\mathfrak{K}\hat{\otimes}\mathfrak{A}^\sim)$, wie sie in (15) betrachtet werden, homotop sind, genau dann, wenn sie konjugiert und damit auch diffeotop sind. Wir können also in (15) die Homotopieklasse $[e]$ durch die Diffeotopieklasse $\langle e\rangle$ ersetzen. Weiter wird in [Ph] gezeigt, dass der Funktor K_*, $* = 0, 1$ auf der Kategorie der Fréchetalgebren die folgenden Eigenschaften hat:

(a) K_* ist diffeotopie- und homotopieinvariant

(b) K_* ist stabil in dem Sinn, dass für jede Fréchetalgebra $\mathfrak{A}$ die Inklusionsabbildung $\mathfrak{A} \to \mathfrak{A}\hat{\otimes}\mathfrak{K}$ einen Isomorphismus in der K-Theorie induziert.

(c) Jede Erweiterung

$$0 \to \mathfrak{J} \xrightarrow{i} \mathfrak{A} \xrightarrow{q} \mathfrak{B} \to 0$$

von Fréchetalgebren (d.h. die Folge ist exakt und q ist eine Quotientenabbildung) induziert exakte Folgen in K_* der folgenden Form:

$$\begin{array}{ccccc} K_0(\mathfrak{J}) & \xrightarrow{K_0(i)} & K_0(\mathfrak{A}) & \xrightarrow{K_0(q)} & K_0(\mathfrak{B}) \\ \uparrow & & & & \downarrow \\ K_1(\mathfrak{B}) & \xleftarrow{K_1(q)} & K_1(\mathfrak{A}) & \xleftarrow{K_1(i)} & K_1(\mathfrak{J}) \end{array}$$

(d) Falls $\mathfrak{A}$ eine Banachalgebra ist, so stimmt $K_*(\mathfrak{A})$ mit der üblichen topologischen K-Theorie von $\mathfrak{A}$ überein.

Für weitere Einzelheiten verweisen wir auf [Ph].

Wir betrachten jetzt die Algebra $Q\mathbb{C}$ und bezeichnen mit $e, \bar{e}$ die beiden Erzeuger $e = \iota(1), \bar{e} = \iota(1))$. Nach 6.3 gilt $\mathbb{Z} \cong K_0(q\mathbb{C}) \subset K_0(Q\mathbb{C}) \cong \mathbb{Z}^2$. Der Erzeuger von $K_0(q\mathbb{C})$ ist mit der oben angegebenen Definition von K_0 für Fréchetalgebren gegeben durch die Diffeotopieklasse des idempotenten Elements p in $M_2(\mathfrak{K}\hat{\otimes}q\mathbb{C}^\sim)$:

$$p = W \begin{pmatrix} \bar{e}^\perp & 0 \\ 0 & e \end{pmatrix} W \qquad \text{wo} \qquad W = \begin{pmatrix} \bar{e}^\perp & \bar{e} \\ \bar{e} & \bar{e}^\perp \end{pmatrix}$$

mit $\bar{e}^\perp = 1 - \bar{e}$. Wir setzen auch

$$\bar{p} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Man beachte, dass $p - \bar{p} \in M_2(\mathfrak{K}\hat{\otimes}q\mathbb{C}) \subset M_2(\mathfrak{K}\hat{\otimes}Q\mathbb{C}^\sim)$ und dass daher $[p] - [\bar{p}] \in K_0(\mathfrak{K}\hat{\otimes}q\mathbb{C}) \subset K_0(\mathfrak{K}\hat{\otimes}Q\mathbb{C})$.

LEMMA 7.1 *Es sei $\varphi : q\mathbb{C} \to M_2(\mathfrak{K}\hat{\otimes}q\mathbb{C})$ die Einschränkung des Homomorphismus $Q\mathbb{C} \to M_2(\mathfrak{K}\hat{\otimes}Q\mathbb{C}^\sim)$, der e auf p und $\bar{e}$ auf $\bar{p}$ abbildet. Dann ist φ diffeotop zu der Inklusionsabbildung $\iota : q\mathbb{C} \to M_2(\mathfrak{K}\hat{\otimes}q\mathbb{C})$.*

Beweis: Sei $\gamma_t : q\mathbb{C} \to M_2(\mathfrak{K}\hat{\otimes}q\mathbb{C}), t \in [0, \pi/2]$ die Einschränkung des Homomorphismus $\gamma'_t : Q\mathbb{C} \to M_2((\mathfrak{K}\hat{\otimes}Q\mathbb{C})^\sim)$ der durch

$$\gamma'_t(e) = W_t \begin{pmatrix} \bar{e}^\perp & o \\ 0 & e \end{pmatrix} W_{-t}$$

$$\gamma'_t(\bar{e}) = W_t \begin{pmatrix} \bar{e}^\perp & 0 \\ 0 & \bar{e} \end{pmatrix} W_{-t}$$

gegeben ist, wo

$$W_t = \begin{pmatrix} \bar{e}^\perp & 0 \\ 0 & \bar{e}^\perp \end{pmatrix} + \begin{pmatrix} \bar{e}\cos t & \bar{e}\sin t \\ -\bar{e}\sin t & \bar{e}\cos t \end{pmatrix}$$

Für jedes t liegt die Differenz $\gamma'_t(e) - \gamma'_t(\bar{e})$ in dem Ideal $M_2(\mathfrak{K}\hat{\otimes}q\mathbb{C}))$. Daher definiert γ_t eine Diffeotopie, die φ mit ι verbindet. *q.e.d.*

SATZ 7.2 *Für jede Fréchetalgebra $\mathfrak{A}$ gilt*

$$K_0(\mathfrak{A}) \cong \langle q\mathbb{C}, \mathfrak{K}\hat{\otimes}\mathfrak{A}\rangle$$

Beweis: Wir definieren die Abbildung $\theta : \langle q\mathbb{C}, \mathfrak{K}\hat{\otimes}\mathfrak{A}\rangle \to K_0(\mathfrak{A})$ in der folgenden Weise: Sei $\eta : q\mathbb{C} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ ein stetiger Homomorphismus. Wir bezeichnen mit z den Erzeuger von $K_0(q\mathbb{C}) \cong \mathbb{Z}$ und setzen $\theta(\langle\eta\rangle) = K_0(\eta)(z)$. Wir zeigen, dass θ surjektiv und injektiv ist.

Die Surjektivität ist offensichtlich, da jedes Element w von $K_0\mathfrak{A}$ nach Definition durch ein idempotentes Element r in $M_2(\mathfrak{K}\hat{\otimes}\mathfrak{A}^\sim)$ gegeben ist und daher einen Homomorphismus $\hat{\eta} : Q\mathbb{C} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}^\sim$ bestimmt, der e auf r und $\bar{e}$ auf $\bar{p}$ ($\bar{p}$ wie oben) abbildet. Die durch $\hat{\eta}$ induzierte Abbildung bildet die durch e und $\bar{e}$ bestimmten Klassen u und $\bar{u}$ in $K_0(Q\mathbb{C})$ auf $[r]$ und $[\bar{p}]$ in $K_0(\mathfrak{A})$ ab. Wenn daher η die Einschränkung von $\hat{\eta}$ auf $q\mathbb{C}$ bezeichnet, so bildet $K_0(\eta)$ den Erzeuger $z = u - \bar{u}$ von $K_0(q\mathbb{C}) \subset K_0(Q\mathbb{C})$ auf $[r] - [\bar{p}] = [r] \in K_0(\mathfrak{A})$ ab.

Um die Injektivität zu beweisen, benutzen wir Lemma 7.1. Nehmen wir an, dass $\eta_1, \eta_2 : q\mathbb{C} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ Homomorphismen sind, so dass $K_0(\eta_1)(z) = K_0(\eta_2)(z)$. Das bedeutet, dass die Bilder r_1 und r_2 des vor 7.1 definierten Idempotenten p unter $M_2(\mathrm{id}_{\mathfrak{K}}\hat{\otimes}\eta_1^\sim)$ und $M_2(\mathrm{id}_{\mathfrak{K}}\hat{\otimes}\eta_2^\sim)$ in $M_2(\mathfrak{K}\hat{\otimes}\mathfrak{A}^\sim)$ konjugiert durch ein invertierbares Element w sind. Dieses Element w kann sogar durch eine differenzierbare Familie $w_t, t \in [0,1]$ mit 1 verbunden werden, so dass $1 - w_t \in M_2(\mathfrak{K}\hat{\otimes}\mathfrak{A})$ für alle t.

Es seien nun $\eta_1', \eta_2' : q\mathbb{C} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ die Homomorphismen $q\mathbb{C} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$, die durch Einschränkung der Abbildungen von $Q\mathbb{C}$, die e auf r_1 bzw. r_2 und $\bar{e}$ auf $\bar{p}$ abbilden, entstehen. Nach Lemma 7.1 ist $\eta_1' = M_2(\mathrm{id}_{\mathfrak{K}}\hat{\otimes}\eta_1^\sim) \circ \varphi$ diffeotop zu $\eta_1 = M_2(\mathrm{id}_{\mathfrak{K}}\hat{\otimes}\eta_1^\sim) \circ \iota$ und η_2' diffeotop zu η_2. Andererseits definiert die Familie $\psi_t, t \in [0,1]$ von Homomorphismen $q\mathbb{C} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$, die durch Einschränkung der Abbildungen von $Q\mathbb{C}$, die e auf $w_t r_1 w_{-t}$ entstehen, eine Diffeotopie, die η_1' mit η_2' verbindet. *q.e.d.*

Für eine beliebige m-Algebra $\mathfrak{A}$ hatten wir in 2.5 die folgende linear zerfallende Erweiterung betrachtet:

$$0 \to q\mathfrak{A}(0,1) \to \mathfrak{E} \to \mathfrak{A} \to 0$$

Wenn wir die klassifizierende Abbildung $J\mathfrak{A} \to q\mathfrak{A}(0,1)$ mit der Toeplitzerweiterung

$$0 \to \mathfrak{K}\hat{\otimes}q\mathfrak{A} \to \mathfrak{T}_0\hat{\otimes}q\mathfrak{A} \to q\mathfrak{A}(0,1) \to 0$$

kombinieren, so erhalten wir eine Abbildung $\varepsilon' : J^2\mathfrak{A} \to \mathfrak{K}\hat{\otimes}q\mathfrak{A}$.

LEMMA 7.3 *Sei $\delta : q\mathfrak{A} \to \mathfrak{A}$ die kanonische Auswertungsabbildung (mit den Bezeichnungen von 1.3 ist δ die Restriktion von $id * 0$). Dann ist die Komposition $(id_{\mathfrak{K}}\hat{\otimes}\delta) \circ \varepsilon'$ diffeotop zu $\varepsilon : J^2\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$.*

Beweis: Dies ergibt sich mit Hilfe von Lemma 3.3 aus dem folgenden kommutativen Diagramm

$$\begin{array}{ccccccc}
0 \longrightarrow & q\mathfrak{A}(0,1) & \longrightarrow & \mathfrak{E} & \longrightarrow & \mathfrak{A} & \longrightarrow 0 \\
 & \downarrow \delta(0,1) & & \downarrow \psi & & \| & \\
0 \longrightarrow & \mathfrak{A}(0,1) & \longrightarrow & \mathfrak{A}[0,1] & \longrightarrow & \mathfrak{A} & \longrightarrow 0
\end{array}$$

wo ψ die Restriktion von $(\mathrm{id} * 0)[0,1] : Q\mathfrak{A}[0,1] \to \mathfrak{A}[0,1]$ auf $\mathfrak{E} \subset Q\mathfrak{A}[0,1]$ ist. *q.e.d.*

THEOREM 7.4 *Für jede Fréchetalgebra $\mathfrak{A}$ sind die Gruppen $kk_*(\mathbb{C},\mathfrak{A})$ und $K_*\mathfrak{A}$ natürlich isomorph.*

Beweis: Wir können annehmen, dass $*=0$. Der Fall $*=1$ ergibt sich durch Ersetzung von $\mathfrak{A}$ durch die Einhängung $\mathfrak{A}(0,1)$.

Die Existenz der gewünschten Abbildung $kk_*(\mathbb{C},\mathfrak{A}) \to K_*\mathfrak{A}$ ergibt sich für $*=0$ als Spezialfall aus 6.5. Um den Isomorphismus zu beweisen, müssen wir aber die Abbildung in systematischer Weise explizit konstruieren.
Nach Korollar 6.5 ist $K_0(\varepsilon) : K_0(J^{2n}\mathbb{C}) \to K_0(\mathbb{C})$ ein Isomorphismus. Aus Satz 7.2 erhalten wir also $\mathbb{Z} \cong K_0(\mathbb{C}) \cong K_0(J^{2n}\mathbb{C}) \cong \langle q\mathbb{C}, \mathfrak{K}\hat{\otimes}J^{2n}\mathbb{C}\rangle$, wobei der zweite Isomorphismus durch ε induziert ist. Sei dann

$$\beta_n : q\mathbb{C} \to \mathfrak{K}\hat{\otimes}J^{2n}\mathbb{C}$$

der bis auf Diffeotopie eindeutig bestimmte Homomorphismus, der dem Erzeuger von $\mathbb{Z}$ unter diesem Isomorphismus entspricht, d.h. $\langle\beta_n\rangle = K_0(\varepsilon^n)^{-1}(1)$.

Andererseits sei $\alpha_n : J^{2n}\mathbb{C} \to \mathfrak{K}\hat{\otimes}q\mathbb{C}$ der Homomorphismus, der sich durch Komposition von $\varepsilon^{n-1} : J^{2n}\mathbb{C} \to \mathfrak{K}\hat{\otimes}J^2\mathbb{C}$ mit der Abbildung $\varepsilon' : \mathfrak{K}\hat{\otimes}J^2\mathbb{C} \to \mathfrak{K}\hat{\otimes}q\mathbb{C}$ aus Lemma 7.3 ergibt.

Lemma 7.3 zeigt dann, dass $(\mathrm{id}_{\mathfrak{K}}\hat{\otimes}\delta)\circ\alpha_n$ diffeotop zu ε^n ist.

Nach Lemma 6.3(b) ist $K_0(\delta) : K_0(q\mathbb{C}) \to K_0\mathbb{C}$ ein Isomorphismus. Da

$$K_0(\mathrm{id}_{\mathfrak{K}}\hat{\otimes}\delta)\circ K_0(\alpha_n)\circ K_0(\beta_n)$$

nach Konstruktion der Isomorphismus $K_0(\delta) : K_0(q\mathbb{C}) \to K_0\mathbb{C}$ ist, folgt daher nach Satz 7.2, dass $\alpha_n\circ\beta_n$ diffeotop zur Inklusion $\iota : q\mathbb{C} \to \mathfrak{K}\hat{\otimes}q\mathbb{C}$ ist und dass $\varepsilon^n\circ\beta_n$ diffeotop zu $\iota\circ\delta : q\mathbb{C} \to \mathfrak{K}$ ist.

Wir können jetzt die Isomorphismen zwischen $kk_0(\mathbb{C},\mathfrak{A})$ und $K_0\mathfrak{A}$ in beide Richtungen explizit angeben. Die Abbildung $\alpha^T : K_0\mathfrak{A} \to kk_0(\mathbb{C},\mathfrak{A})$ bildet $\langle\gamma\rangle \in \langle q\mathbb{C}, \mathfrak{K}\hat{\otimes}\mathfrak{A}\rangle$ auf die Klasse von $(\mathrm{id}_{\mathfrak{K}}\otimes\gamma)\circ\alpha_n : J^{2n}\mathbb{C} \to \mathfrak{K}\hat{\otimes}A$ in $kk_0(\mathbb{C},\mathfrak{A})$ ab. Die umgekehrte Abbildung $\beta^T : kk_0(\mathbb{C},\mathfrak{A}) \to K_0\mathfrak{A}$ ist folgendermaßen definiert: Sei $\eta : J^{2n}\mathbb{C} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ ein Repräsentant für ein Element h in $kk_0(\mathbb{C},\mathfrak{A})$. Wir setzen dann $\beta^T(h) = \langle\eta\circ\beta_n\rangle \in \langle q\mathbb{C}, \mathfrak{K}\hat{\otimes}\mathfrak{A}\rangle$. Nach Konstruktion von β_n hängt diese Diffeotopieklasse nicht von der Auswahl des Repräsentanten η ab, und $\beta^T(h)$ ist daher wohldefiniert. Aus der obigen Diskussion folgt sofort, dass $\beta^T\circ\alpha^T = \mathrm{id}$.

Um die Komposition $\alpha^T\circ\beta^T$ zu berechnen. benutzen wir wieder das Hauptlemma 3.10 und sein Korollar 3.11, d.h. im Grund das Produkt in kk. Sei h ein Element von $kk_0(\mathbb{C},\mathfrak{A})$, das durch einen Homomorphismus $\eta : J^{2n}\mathbb{C} \to \mathfrak{K}\hat{\otimes}\mathfrak{A}$ repräsentiert ist. Nach Korollar 3.11 sind die folgenden beiden Kompositionen diffeotop

$$(\mathrm{id}_{\mathfrak{K}}\otimes\eta)\circ\varepsilon^n\circ J^{2n}((\mathrm{id}_{\mathfrak{K}}\otimes\beta_n)\circ\alpha_n) \sim (\mathrm{id}_{\mathfrak{K}}\otimes\eta)((\mathrm{id}_{\mathfrak{K}}\otimes\varepsilon^n)\circ\beta_n\circ\alpha_n)_{2n}$$

(unter Verwendung der Bezeichnungsweise $\psi_j : J^{j+k}\mathfrak{A} \to \mathfrak{K}\hat{\otimes}J^j\mathfrak{B}$ für $\psi : J^k\mathfrak{A} \to \mathfrak{K}\hat{\otimes}\mathfrak{B}$, die vor Lemma 4.2 eingeführt wurde). Da $\varepsilon^n\circ J^{2n}(\varphi) = \varphi\circ\varepsilon^n$ für alle φ, repräsentiert die erste Komposition $\alpha^T\circ\beta^T(h)$. Da andererseits $\varepsilon^n : J^{2n}(J^{2n}\mathbb{C}) \to \mathfrak{K}\hat{\otimes}J^{2n}\mathbb{C}$, wieder nach Korollar 3.11, diffeotop zu $(\varepsilon^n)_n$ ist und weil $\varepsilon^n\circ\beta_n\circ\alpha_n \sim \varepsilon^n$, repräsentiert die zweite Komposition gerade h. Damit ist gezeigt, dass $\alpha^T\circ\beta^T = \mathrm{id}$. *q.e.d.*

Wie schon in der Einleitung erwähnt, kann Theorem 7.4 für abgeleitete Unteralgebren von Banachalgebren (im Sinn von 1.6) ohne Verwendung der Theorie von Phillips direkt bewiesen werden. Wir skizzieren kurz, wie man vorzugehen hat.

Die Topologie auf $J^{2n}\mathbb{C}$ ist gegeben durch die Familie von submultiplikativen Normen, die auf $T^{2n}\mathbb{C}$ durch die Vielfachen der kanonischen Norm auf $\mathbb{C}$ induziert werden. Jeder stetige Homomorphismus φ von $J^{2n}\mathbb{C}$ in eine abgeleitete Unteralgebra einer Banachalgebra A ist stetig für eine dieser Normen und setzt sich daher auf die entsprechende Vervollständigung B von $J^{2n}\mathbb{C}$ fort. Nach Lemma 1.6.5 ist φ dann auf einer abgeleiteten Unteralgebra $\mathfrak{B}$ von B definiert. Die K-Theorie dieser abgeleiteten Unteralgebra ist wohldefiniert und stimmt mit der von B überein.

Insbesondere kann das auf den stetigen Homomorphismus $\varepsilon : J^{2n}\mathbb{C} \to \mathfrak{K}$ angewendet werden und wie in Korollar 6.5 sieht man sofort, dass ε einen Isomorphismus $K_0(\mathfrak{B}) \to K_0(\mathfrak{K})$ induziert.

Der Beweis von Satz 7.4 benutzt nur die Definition der K-Theorie durch Diffeotopieklassen von Idempotenten in $M_2(\mathfrak{K}\hat{\otimes}\mathfrak{A})^{\sim}$, die für abgeleitete Unteralgebren in derselben Form gilt.

Schließlich können Homomorphismen von $q\mathbb{C}$ ebenso behandelt werden wie die von $J^{2n}\mathbb{C}$ und auf abgeleitete Unteralgebren von Banachalgebravervollständigungen fortgesetzt werden. Die Homomorphismen α_n und β_n im Beweis zu Theorem 7.4 können dann als Homomorphismen zwischen solchen Vervollständigungen (die aber von dem gegebenen Homomorphismus η abhängen) konstruiert werden.

8 Vergleich der Filtrierungen in kk und HP.

Für beliebige m-Algebren $\mathfrak{A}$ und $\mathfrak{B}$ gilt $kk_0(\mathfrak{A},\mathfrak{B}) = kk_0(\mathfrak{A}, J^2\mathfrak{B})$, vgl. 4.5, und $\langle\mathfrak{A}, \mathfrak{K}\hat{\otimes}\mathfrak{B}\rangle \cong \langle\mathfrak{K}\hat{\otimes}\mathfrak{A}, \mathfrak{K}\hat{\otimes}\mathfrak{B}\rangle$ (als Konsequenz aus 1.4.1). Hieraus ergibt sich die folgende alternative Definition von kk_0

$$kk_0(\mathfrak{A},\mathfrak{B}) = \varprojlim_m \left(\varinjlim_n \langle\mathfrak{K}\hat{\otimes}J^{2n}\mathfrak{A}, \mathfrak{K}\hat{\otimes}J^{2m}\mathfrak{B}\rangle\right)$$

Damit erhalten wir eine sehr einfache Beschreibung des Produkts in kk_0, das nämlich genau wie das Produkt von Morphismen zwischen Pro-Objekten definiert ist. Die Wohldefiniertheit und Assoziativität des Produkts ist dann völlig offensichtlich.
Die obige Beschreibung von kk_0 ist nun aber auch formal fast genau analog zur Definition der bivarianten periodischen zyklischen Homologie. Wir erinnern daran, dass diese in der folgenden Weise definiert werden kann

$$HP^*(\mathfrak{A},\mathfrak{B})) = H_*\Big(\varprojlim_m \left(\varinjlim_n Hom(X(J\mathfrak{A})^n, X(J\mathfrak{B})^m)\right)\Big)$$

siehe [CuQu2,3.2]. Der wichtigste Unterschied in den Formeln für kk_0 und HP^* ist die Tatsache, dass einmal die durch Iteration des J-Funktors erhaltenen Algebren $J^n\mathfrak{A}$ und $J^m\mathfrak{B}$ benutzt werden und das andere Mal die Potenzen $(J\mathfrak{A})^n$ und $(J\mathfrak{B})^m$.
Wenden wir uns jetzt wieder der Definition von $kk_*(\mathfrak{A},\mathfrak{B})$, wie sie in 4.1 gegeben wurde, zu. Diese führt unmittelbar zu einer natürlichen aufsteigenden Filtrierung

durch die Bilder von $\langle J^{2n+*}\mathfrak{A}, \mathfrak{K}\hat{\otimes}\mathfrak{B}\rangle$ in

$$kk_*(\mathfrak{A}, \mathfrak{B}) = \varinjlim_n \langle J^{2n+*}\mathfrak{A}, \mathfrak{K}\hat{\otimes}\mathfrak{B}\rangle$$

Auf der anderen Seite besitzt die periodische zyklische Kohomologie $HP^*(\mathfrak{A})$ und das Bild der bivarianten Jones-Kassel Theorie in $HP^*(\mathfrak{A}, \mathfrak{B})$ eine natürliche Filtrierung durch die Bilder von HC^n. Da die Filtrierungen von kk_* und HP^* beide mit dem Produkt verträglich sind, und der Chern-Connes-Charakter multiplikativ ist, werden die Filtrierungen unter dem Charakter wenigstens teilweise erhalten. Man kann etwa eine Unterhalbgruppe $ext_*(\mathfrak{A}, \mathfrak{B})$ von $kk_*(\mathfrak{A}, \mathfrak{B})$ einführen, die aus allen Yonedaprodukten von Erweiterungen von dem Typ, wie sie in [Ni1] betrachtet werden, besteht. Diese Unterhalbgruppe trägt eine natürliche Filtrierung. Die Konstruktion aus [Ni1] zeigt, dass die Filtrierung unter dem Chern-Connes-Charakter erhalten wird.
Für beliebige Element von kk andererseits zeigt Satz 6.12 durch Iteration, dass eine natürliche Abbildung $J^{2p+1}\mathfrak{A} \to \mathfrak{K}\hat{\otimes}(J\mathfrak{A})^{2^p}$ existiert. Dies legt nahe, dass im allgemeinen in gewissem Sinn die Ordnung der Filtrierung auf kk dem Logarithmus der Ordnung der Filtrierung auf HP^*, d.h. dem Logarithmus der Dimension entspricht. Eine genauere Untersuchung bleibt einer weiteren Arbeit vorbehalten.
Als letztes bemerken wir, dass auch bei der Definition der K-Theorie noch interessante Variationen möglich sind. Wir können etwa setzen

$$k_n(\mathfrak{A}) = \varinjlim_k \langle J^{k-n}\mathbb{C}, \mathfrak{K}\hat{\otimes} J^k\mathfrak{A}\rangle$$

Ein Argument wie im Beweis zu Theorem 7.4 zeigt, dass für $n \geq 1$ jeweils

$$\langle J^{2k+2n}\mathbb{C}, \mathfrak{K}\hat{\otimes} J^{2k}\mathfrak{A}\rangle = \langle q\mathbb{C}, \mathfrak{K}\hat{\otimes} J^{2k}\mathfrak{A}\rangle = K_{2n}(\mathfrak{A})$$

Für negative n ist also k_n periodisch und stimmt mit der K-Theorie überein. Für positive n ergibt sich eine Art konnektiver K-Theorie, vgl. [Se], [Ro] mit einer Periodizitätsabbildung $k_n(\mathfrak{A}) \to k_{n-2}(\mathfrak{A})$, die durch ε induziert wird.

LITERATUR

[Bl] B.Blackadar: K-theory for Operator Algebras, Springer-Verlag, Heidelberg / Berlin / New York / Tokyo, 1986.

[BlCu] B. Blackadar und J. Cuntz: Differential Banach algebra norms and smooth subalgebras of C^*-algebras, J. Operator Theory 26 (1991), 255-282.

[Bo] J.-B.Bost: Principe d'Oka, K-théorie et systèmes dynamiques non-commutatifs, Invent. Math. 101, 261 - 333 (1990).

[BDF] L.G.Brown, R.G.Douglas und P.Fillmore: Extensions of C*-algebras and K-homology, Ann. of Math. 105 (1977), 265-324.

[Co1] A. Connes: Non-commutative differential geometry. Publ. Math. I.H.E.S. 62 (1985), 257-360.

[Co2] A. Connes: Non-commutative Geometry, Academic Press, London-Sydney-Tokyo-Toronto, 1994.

[CoHi] A. Connes und N.Higson: Déformations, morphismes asymptotiques et K-théorie bivariante, C.R.Acad.Sci. Paris Ser. I, Math, 311 (1990), 101-106.

[CoMo] A. Connes und H. Moscovici: Cyclic cohomology, the Novikov conjecture and hyperbolic groups, Topology, 29 (1990), 345-388.

[Cu1] J. Cuntz: K-theory and C*-algebras, in "Algebraic K-theory, Number theory, Geometry and Analysis", Springer Lecture Notes in Math. 1046, 55-79.

[Cu2] J. Cuntz: A new look at KK-theory, K-theory 1 (1987), 31-52.

[Cu3] J. Cuntz: A survey of some aspects of non-commutative geometry, Jber. d. Dt. Math.-Verein. 95 (1993) 60-84.

[Cu4] J. Cuntz: Excision in bivariant cyclic theory for topological algebras, Proc. Workshop on Cyclic Cohomology and Noncommutative Geometry, Fields Institute, wird erscheinen.

[CuQu1] J. Cuntz und D. Quillen: Cyclic homology and nonsingularity, J. Amer. Math. Soc. 8 (1995), 373-442.

[CuQu2] J. Cuntz und D. Quillen: Excision in bivariant periodic cyclic cohomology, Invent. math. 127 (1997), 67-98.

[DoVo] R.G.Douglas und D.Voiculescu: On the smoothness of sphere extensions, J.Operator theory 6 (1981), 103-111.

[ENN] G.A.Elliott, T.Natsume und R.Nest: Cyclic cohomology for one-parameter smooth crossed products, Acta Math. 160 (1988), 285-305.

[Ga] Gauvrit: Cyclic cohomology for Schatten ideals, preprint.

[Gr] A.Grothendieck: Produits tensoriels topologiques et espaces nucléaires, Memoirs of the Amer. Math. Soc. 16 (1955).

[Hi] N.Higson: On Kasparov theory, M.A. thesis, Dalhousie University, 1983.

[JoKa] J.D.S.Jones und C.Kassel: Bivariant cyclic theory, K-theory 3 (1989), 339-366.

[Ka] G. G. Kasparov: The operator K-functor and extensions of C^*-algebras, Izv. Akad. Nauk. SSSR, Ser. Mat. 44 (1980), 571-636

[Ks] C.Kassel: Caractère de Chern bivariant, K-theory 3, 367 - 400 (1989)

[La] V.Lafforgue: private Mitteilung

[Lo] J.-L.Loday: Cyclic Homology, Grundlehren 301, Springer Verlag, 1992.

[Mi] E. A. Michael: Locally multiplicatively-convex topological algebras, Memoirs of the AMS, number 11, 1952.

[Ni1] V.Nistor: A bivariant Chern-Connes character, Ann. of Math. 138, 555 - 590 (1993).

[Ni2] V.Nistor: On the Cuntz-Quillen boundary map, preprint 1994 / Higher index theorems and the boundary map in cyclic cohomology, preprint 1997.

[Ph] C.Phillips: K-theory for Fréchet algebras, International Journal of Mathematics vol. 2 no. 1 (1991), 77 - 129.

[Ro] J.Rosenberg: The role of K-theory in noncommutative algebraic topology, Operator algebras and K-theory, Contemp. Math. 10, Amer. math. Soc. 1982, 155 - 182.

[Se] G.Segal: K-homology theory and algebraic K-theory, K-theory and Operator algebras, Lecture Notes in Math. 575, 113 - 127.

[Sk] G.Skandalis: Une notion de nucléarité en K-théorie, K-theory 1 (1988), 549-573.

[T] F.Treves: Topological vector spaces, distributions and kernels, Academic Press, New York, London, 1967.

[Pu] M.Puschnigg: Asymptotic cyclic cohomology, Lecture Notes in Mathematics 1642, Springer 1996.

[Wa] L. Waelbroeck: Topological vector spaces and algebras, Lecture Notes in Math. 230, Springer-Verlag; 1971.

[Wa] X.Wang: A bivariant Chern character, Can. J. Math. 35, 1 - 36 (1992).

[We], J.Weidner: KK-groups for generalized operator algebras I, K-theory 3 (1989), 57-77.

[Wo] M.Wodzicki: Excision in cyclic homology and in rational algebraic K-theory, Ann. of Math. 129 (1989), 591-639.

[Ze] R.Zekri: A new description of Kasparov's theory of C*-algebra extensions, J. Functional Analysis 84 (1989), 441 - 471.

Joachim Cuntz
Mathematisches Institut
der Universität
Im Neuenheimer Feld 288
69120 Heidelberg
Germany
cuntz@math.uni-heidelberg.de

Compact Complex Manifolds with Numerically Effective Cotangent Bundles

Henrik Kratz

Received: June 23, 1997

Communicated by Thomas Peternell

Abstract. We prove that a projective manifold of dimension $n = 2$ or 3 and Kodaira dimension 1 has a numerically effective cotangent bundle if and only if the Iitaka fibration is almost smooth, i.e. the only singular fibres are multiples of smooth elliptic curves ($n = 2$) resp. multiples of smooth Abelian or hyperelliptic surfaces ($n = 3$). In the case of a threefold which is fibred over a rational curve the proof needs an extra assumption concerning the multiplicities of the singular fibres. Furthermore, we prove the following theorem: let X be a complex manifold which is hyberbolic with respect to the Carathéodory-Reiffen-pseudometric, then any compact quotient of X has a numerically effective cotangent bundle.

1991 Mathematics Subject Classification: 32C10, 32H20

Introduction

It is a natural question in algebraic geometry to classify manifolds by positivity properties of their tangent resp. cotangent bundles. The first result of this kind was obtained by Mori who solved the Hartshorne-Frankel conjecture [Mo]: every projective n-dimensional manifold with ample tangent bundle is isomorphic to the complex projective space $\mathbb{P}_n$. A degenerate condition of ampleness is numerical effectivity. A line bundle L on a projective manifold X is called numerically effective (abbreviated "nef") if $L.C \geq 0$ for all curves $C \subset X$. A vector bundle E is said to be nef if the tautological quotient line bundle $\mathcal{O}_{\mathbb{P}(E)}(1)$ on $\mathbb{P}(E)$, the projective bundle of hyperplanes in the fibres of E, is nef.

Taking the Hartshorne-Frankel conjecture as a guideline, Campana and Peternell considered projective manifolds whose tangent bundles are nef and classified them in dimension 2 and 3 [CP]. For dimension 3 this has been done by Zheng [Zh] too. In general, for arbitrary compact complex manifolds the "nefness" of the tangent bundle leads to strong structural constraints [DPS].

The purpose of this paper is to investigate some aspects of manifolds X whose cotangent bundles Ω^1_X are nef. In the first part we will give a characterization of 2 and 3 dimensional manifolds with Kodaira dimension $\kappa(X) = 1$ and nef cotangent bundle. We will prove:

Theorem 1 *Let X be a minimal projective manifold of dimension $n = 2$ or 3 with $\kappa(X) = 1$ and let $\pi : X \to C$ be the Iitaka fibration of X. Then the following conditions are equivalent:*
(i) Ω^1_X *is nef.*
(ii) π *is almost smooth, in the sense that the only singular fibres of π are multiples of smooth elliptic curves $(n = 2)$ resp. Abelian or hyperelliptic surfaces $(n = 3)$.*
• *Exception: To prove* (ii)$\Rightarrow$(i) *in the case $n = 3$ and $g(C) = 0$ we need the assumption that $\sum \frac{m_i - 1}{m_i} \geq 2$, where the m_i are the multiplicities of the singular fibres.*
• *The equivalence of* (i) *and* (ii) *holds also for compact Kähler surfaces.*

This theorem generalizes a result of Fujiwara [Fu] who worked in arbitrary dimension but under the stronger assumption that Ω^1_X is semi-ample, i.e. that some power of $\mathcal{O}_{\mathbb{P}(\Omega^1_X)}(1)$ is globally generated. The implication (i) $\Rightarrow$ (ii) relies on the topological constraints, namely the Chern class inequalities, which hold, when the cotangent bundle is nef. To prove (ii) $\Rightarrow$ (i) we will proceed in two steps. First, we will show that the assertion is true for a smooth fibration. This follows basically from Griffiths's theory on the variation of the Hodge structure. Then, we will study the base-change which reduces an almost smooth fibration to a smooth one and show that this process allows to carry over the "nefness" of the cotangent bundle.
In fact, we will prove in any dimension that a projective manifold has a nef cotangent bundle if (a) it admits a smooth Abelian fibration over a manifold with nef cotangent bundle or (b) it admits an almost smooth Abelian fibration over a curve C such that either (i) $g(C) \geq 1$ or (ii) $g(C) = 0$ and $\sum \frac{m_i - 1}{m_i} \geq 2$.
We remark that the fibres F of the Iitaka fibrations in Theorem 1 are paraAbelian varieties, i.e. there exists an unramified cover $T \to F$ where T is an Abelian variety. In view of this, we expect in any dimension that a manifold with Kodaira dimension 1 has a nef cotangent bundle if and only if the Iitaka fibration is almost smooth with para-Abelian fibres.
In the second part of this paper we consider complex manifolds X which are hyperbolic with respect to the Carathéodory-Reiffen pseudometric. We will show :

Theorem 2 *Let X be a complex manifold which is hyperbolic with respect to the Carathéodory-Reiffen pseudometric and let Q be a compact quotient of X with respect to a subgroup of the automorphism group of X which operates fixpointfree and properly discontinuously. Then Ω^1_Q is nef.*

In particular, any compact quotient of a bounded domain $G \subset \mathbb{C}^n$ possesses a nef cotangent bundle. Since the canonical bundle of such a quotient is ample, this yields a class of manifolds with maximal Kodaira dimension and nef cotangent bundle.
To prove theorem 2 we apply the technique of singular hermitian metrics which was developed by Demailly. The Carathéodory-Reiffen pseudometric of X defines a Finsler structure on the tangent bundle of Q and this gives us a singular hermitian metric on $\mathcal{O}_{\mathbb{P}(\Omega^1_Q)}(1)$. The hyperbolicity of X guarantees that this metric is continuous and that the associated curvature current is positive. These conditions imply that $\mathcal{O}_{\mathbb{P}(\Omega^1_Q)}(1)$ is nef.

Acknowledgments: I would like to thank M. Schneider and Th. Peternell for their help and encouragement.

1 BASIC DEFINITIONS AND PROPERTIES

Let X and Y be compact complex manifolds and let L be a holomorphic line bundle on X.

DEFINITION 1 (i) *When X is projective, L is said to be nef, if $L \cdot C = \int_C c_1(L) \geq 0$ for every curve C in X.*
(ii) *Let X be an arbitrary compact complex manifold equipped with a hermitian metric ω. Then L is said to be nef, if for all $\epsilon > 0$ there exists a smooth hermitian metric h_ϵ on L such that the associated curvature form satisfies*

$$\Omega_{h_\epsilon}(L) \geq -\epsilon \cdot \omega.$$

(iii) *Let E be a holomorphic vector bundle on X and $\mathbb{P}(E)$ the projective bundle of hyperplanes in the fibres of E. Then we call E nef over X, if the tautological quotient line bundle $\mathcal{O}_{\mathbb{P}(E)}(1)$ is nef over $\mathbb{P}(E)$.*

We will frequently use the following propositions which are proved in [DPS].

PROPOSITION 1 *Let $f : Y \to X$ be a holomorphic map and let E be a holomorphic vector bundle over X. Then E nef implies f^*E nef, and the converse is true if f is surjective and has equidimensional fibres.*

PROPOSITION 2 *Let E and F be holomorphic vector bundles. Then*
(i) *E, F nef $\Rightarrow E \otimes F$ nef.*
(ii) *E nef $\Rightarrow \det(E)$ nef.*

PROPOSITION 3 *Let $0 \to F \to E \to Q \to 0$ be an exact sequence of holomorphic vector bundles. Then*
(i) *E nef $\Rightarrow Q$ nef.*
(ii) *F, Q nef $\Rightarrow E$ nef.*

Proposition 1 immediately implies

PROPOSITION 4 *Let Y be a finite unramified covering of X. Then Ω^1_X is nef if and only if Ω^1_Y is nef.*

A fibration of X over Y is a surjective holomorphic map $\pi : X \to Y$ whose fibres are connected. A point $x \in X$ is said to be critical if the tangent map $D\pi(x)$ has not maximal rank. The images $\pi(x) \in Y$ of the critical points are the critical values of π. They form a proper analytic subset of Y, i.e. in the case, where Y is a curve, a finite subset $\{a_1, \dots, a_l\}$.
Let $y \in Y$ and let $\mathcal{J}$ be the ideal sheaf of y in $\mathcal{O}_Y$. Then the fibre X_y is the complex subspace $(\pi^{-1}(y), \mathcal{O}_X/\pi^*(\mathcal{J}) \cdot \mathcal{O}_X)$ of X, and a fibre X_y is singular if and only if y is a critical value. A fibration, for which $D\pi$ has maximal rank everywhere, is called smooth.
When we consider a fibration $\pi : X \to C$ over a curve C, we will always assume that C is smooth. Such a fibration is said to be almost smooth, if the only singular fibres of π are multiples of smooth irreducible subvarieties. Their multiplicities will be denoted by m_i with $1 \leq i \leq l$, so that the singular fibres are $X_{a_i} = m_i F_i$, where the F_i are smooth irreducible subvarieties.

We will denote the Kodaira dimension of X by $\kappa(X)$. Let X be a projective manifold with $\kappa(X) \geq 1$ for which a power of the canonical bundle is globally generated. Then for m big enough the $m-$canonical map gives us a holomorphic map $\pi : X \to Z$ where Z is a projective variety with $\dim Z = \kappa(X)$. Such a map π is called Iitaka fibration (cf. [Ue]).

2 MANIFOLDS WITH $\kappa = 1$ AND NEF COTANGENT BUNDLE

We will now prove

THEOREM 3 *Let X be a minimal projective manifold of dimension $n = 2$ or 3 with $\kappa(X) = 1$ and let $\pi : X \to C$ be the Iitaka fibration of X. Then the following conditions are equivalent:*
(i) *Ω^1_X is nef.*
(ii) *π is almost smooth, in the sense that the only singular fibres of π are multiples of smooth elliptic curves $(n = 2)$ resp. Abelian or hyperelliptic surfaces $(n = 3)$.*
• *Exception: To prove* (ii)$\Rightarrow$(i) *in the case $n = 3$ and $g(C) = 0$ we need the assumption that $\sum \frac{m_i - 1}{m_i} \geq 2$, where the m_i are the multiplicities of the singular fibres.*
• *The equivalence of* (i) *and* (ii) *holds also for compact Kähler surfaces.*

Proof: (i) $\Rightarrow$ (ii) If X is an n-dimensional projective manifold with Ω^1_X nef, it satisfies the Chern class inequality $c_1(X)^2 \geq c_2(X) \geq 0$, i.e.

$$c_1(X)^2 \cdot H_1 \cdot \ldots \cdot H_{n-2} \geq c_2(X) \cdot H_1 \cdot \ldots \cdot H_{n-2} \geq 0$$

for all ample divisors H_i (cf. [DPS], Thm. 2.5). For $n = 2$ and 3 the abundance conjecture holds which means that a power of the canonical bundle of X has to be globally generated so that we get from $\kappa(X) = 1$ that $c_1(X)^2 \equiv 0$ and hence $c_1(X)^2 \equiv c_2(X) \equiv 0$. Here $\equiv$ denotes numerical equivalence.
So for $n = 2$ we have an elliptic surface X whose topological Euler characteristic is $e(X) = c_2(X) = 0$. On the other hand, if $\pi : X \to C$ is the Iitaka fibration of X and X_{a_i} are the singular fibres $(1 \leq i \leq l)$, we calculate $e(X) = \sum e(X_{a_i})$. But now the assertion follows, because $e(X_{a_i}) \geq 0$ and $e(X_{a_i}) = 0$ if and only if the fibre X_{a_i} is a multiple of a smooth elliptic curve (cf. [BPV], Chap. III, Prop. 11.4). This argument remains true for a compact Kähler surface.
For $n = 3$ we have a minimal threefold with the extremal Chern classes $c_1(X)^2 \equiv 3c_2(X) \equiv 0$ and the assertion follows from [PW], Theorem 2.1.
(ii) $\Rightarrow$ (i) We will prove this direction by reducing it to the case of a smooth fibration.

2.1 SMOOTH FIBRATIONS

We will consider smooth Abelian fibrations first:

PROPOSITION 5 *Let X and Y be projective manifolds and let $\pi : X \to Y$ be a smooth fibration, whose fibres are Abelian varieties. Then the relative cotangent bundle $\Omega^1_{X/Y}$ is nef. If Ω^1_Y is nef, Ω^1_X is nef too.*

Proof: (1) We claim that $\pi^*(\pi_*\Omega^1_{X/Y}) = \Omega^1_{X/Y}$. For all $y \in Y$ the cotangent bundle of the fibre $\Omega^1_{X_y}$ is trivial, so that $\pi_*\Omega^1_{X/Y}$ is locally free of rank equal to the dimension

of the fibres (cf. [Ha], Chap. III, Cor. 12.9). Moreover for all $y \in Y$ we have $(\pi_*\Omega^1_{X/Y})_y \cong H^0(X_y, \Omega^1_{X_y})$ and thus $(\pi^*(\pi_*\Omega^1_{X/Y}))_x \cong H^0(X_y, \Omega^1_{X_y})$ for $\pi(x) = y$. Now, the canonical homomorphism $\alpha : \pi^*(\pi_*\Omega^1_{X/Y}) \to \Omega^1_{X/Y}$ is described stalkwise by $\alpha_x : \sigma \mapsto \sigma(x)$ with $\sigma \in H^0(X_y, \Omega^1_{X_y})$. Since $\Omega^1_{X/Y} \mid_{X_y}$ is globally generated, α_x is surjective and hence bijective.
(2) Any smooth fibration $\pi : X \to Y$ of projective manifolds gives rise to a variation of the Hodge structure in its fibres X_y $(y \in Y)$. From this Griffiths deduces [Gr], Cor. 7.8

THEOREM 4 *For all $n \in \{1, \ldots, \dim_{\mathbb{C}} X_y\}$ the bundles $R^n\pi_*(\mathcal{O}_X)$ are seminegative in the sense of Griffiths.*

Now the bundle $E = R^n\pi_*(\mathcal{O}_X)$ is conjugate linear to $\bar{E} = \pi_*(\Omega^n_{X/Y})$ so that the curvature matrices with respect to unitary bases behave as

$$\Omega_{\bar{E}} = \bar{\Omega}_E = -\Omega^t_E.$$

Since the transposition of the curvature matrix does not change its positivity properties, the preceding theorem can equivalently be formulated as

THEOREM 5 *For all $n \in \{1, \ldots, \dim_{\mathbb{C}} X_y\}$ the bundles $\pi_*(\Omega^n_{X/Y})$ are semipositive in the sense of Griffiths.*

In particular, since semipositivity implies "nefness", $\pi_*(\Omega^n_{X/Y})$ is nef and hence for a smooth Abelian fibration $\Omega^1_{X/Y} = \pi^*(\pi_*\Omega^1_{X/Y})$ is nef too. The second assertion follows immediately from the relative cotangent sequence and Proposition 3.

Remark: Proposition 5 holds also for compact elliptic surfaces $\pi : X \to C$, because for a smooth π one knows from the study of the period map that $\deg(\pi_*\omega_{X/C}) = 0$ (cf. [BPV], Chap. III, Thm. 18.2).
We have a similar proposition for smooth hyperelliptic fibrations:

PROPOSITION 6 *Let X be a projective 3-dimensional manifold and let $\pi : X \to C$ be a smooth fibration, whose fibres are hyperelliptic surfaces. Furthermore, let $g(C) \geq 1$. Then Ω^1_X is nef.*

Proof: We consider the relative Albanese factorization of π, i.e. the commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{A_\pi} & A(X/C) \\ & {\scriptstyle \pi} \searrow & \downarrow {\scriptstyle Alb(\pi)} \\ & & C, \end{array}$$

where $A(X/C)$ is a smooth fibration over C whose fibres over $a \in C$ are the Albanese tori $\mathrm{Alb}(X_a)$ of the fibres X_a of π. The existence of such a relative Albanese diagram is proved in [Ca]. Since the tangent bundle of a hyperelliptic surface is nef, the Albanese map $A_\pi \mid_{X_a}: X_a \to \mathrm{Alb}(X_a)$ is a surjective submersion with smooth elliptic curves as fibres ([DPS], Prop. 3.9.). But also A_π is smooth: let $x \in X, \pi(x) = a$ and $A_\pi(x) = y$, then both tangent directions of $TA(X/C)_y$ lie in the image of $DA_\pi(x)$. First, we can find a tangent vector $v \in (TA(X/Y) \mid_{\mathrm{Alb}(X_a)})_y$ in the image of $DA_\pi(x) \mid_{X_a}$ (because $A_\pi \mid_{X_a}$ is smooth). Now let (x_1, x_2, x_3) be a coordinate system centered in x and let z_1 be a coordinate centered in a, such that $D\pi(x).\frac{\partial}{\partial x_1} = \frac{\partial}{\partial z_1}$. Using the commutativity of the relative Albanese diagram, we get

$$0 \neq D\pi(x).\frac{\partial}{\partial x_1} = DAlb(\pi)(y) \circ DA_\pi(x).\frac{\partial}{\partial x_1}.$$

In particular, $w := DA_\pi(x).\frac{\partial}{\partial x_1} \neq 0$, and since $DAlb(\pi)(y).v = 0$ the vectors v and w have to be linear independent.
We can now apply Proposition 5 twice to conclude that Ω^1_X is nef: $Alb(\pi) : A(X/C) \to C$ is a smooth fibration of projective manifolds whose fibres are elliptic curves and by assumption $g(C) \geq 1$, so that $\Omega^1_{A(X/C)}$ has to be nef. Since $A_\pi : X \to A(X/C)$ is a smooth elliptic fibration too, Ω^1_X is also nef.

2.2 Almost smooth fibrations

Let X be a compact complex manifold of dimension n and let $\pi : X \to C$ be an almost smooth fibration over a smooth curve C. As above we will denote the critical values of π by $a_1, \ldots, a_l$ and their multiplicities bym_i where $1 \leq i \leq l$, so that the singular fibres are $X_{a_i} = m_i F_i$, where the F_i are smooth irreducible subvarieties.
To get rid of the multiple fibres we will now perform a base change which was introduced by Kodaira for elliptic surfaces ([Kod], Thm 6.3), but may be used in this general context as well. Let m_0 be the lowest common multiple of the multiplicities and let d be their product. Then we choose a finite covering $\sigma : C' \to C$, which has $\frac{d}{m_i}$ ramification points of order $m_i - 1$ over the points a_i where $0 \leq i \leq l$. Remark that we have to add one extra point a_0 which is not contained in the set of critical values. Then the normalization of the fibre product $X \times_C C'$ gives us a smooth fibration $\varphi : X' \to C'$ and a commutative diagram (cf. [Kod], Thm 6.3)

$$\begin{array}{ccc} X' & \xrightarrow{f} & X \\ \varphi \downarrow & & \downarrow \pi \\ C' & \xrightarrow{\sigma} & C \end{array} \quad .$$

Here f is a finite covering which is unramified over $X - \pi^{-1}(a_0)$, because the multiplicities of π and σ compensate each other over a_i $(i \geq 1)$, and f has $\frac{d}{m_0}$ ramification divisors of order $m_0 - 1$ over $\pi^{-1}(a_0)$.
Assume that we knew $\Omega^1_{X'}$ is nef, then we would like to carry this over to Ω^1_X. However, it is not possible to apply Proposition 4 since f is ramified. But we have the following commutative diagram with exact rows which was already used in [Fu]

$$\begin{array}{ccccccc} 0 \longrightarrow & f^*(L) & \longrightarrow & f^*(\Omega^1_X) & \longrightarrow & \Omega^1_{X'/C'} & \longrightarrow 0 \\ & \downarrow & & \downarrow & & \| & \\ 0 \longrightarrow & \varphi^*(K_{C'}) & \longrightarrow & \Omega^1_{X'} & \longrightarrow & \Omega^1_{X'/C'} & \longrightarrow 0. \end{array}$$

Let $D = \sum_{i=1}^{l}(m_i - 1)F_i$ then $L = \pi^*(K_C) \otimes \mathcal{O}_X(D)$ is the full subbundle of Ω^1_X associated to $\pi^*(K_C)$ (cf. [Re]). To prove the commutativity of this diagram one uses basically the fact that the restriction of f to a fibre of φ is unramified. For $i \geq 1$ we have $\pi^*(a_i) = m_i F_i$. So, defining $A := \sum_{i=1}^{l} \frac{(m_i-1)}{m_i} \cdot a_i$ we get $L = \pi^*(K_C \otimes \mathcal{O}_C(A))$. Combining the diagram and Proposition 5, we obtain

COROLLARY 1 *Let X be a projective manifold of arbitrary dimension and let $\pi : X \to C$ be an almost smooth fibration, whose fibres are Abelian varieties. Assume furthermore that* (i) $g(C) \geq 1$ *or* (ii) $g(C) = 0$ *and* $\deg A \geq 2$. *Then Ω^1_X is nef.*

Proof: The process described above allows us to pass to a smooth Abelian fibration φ, for which $\Omega^1_{X'/C'}$ is nef by Proposition 5. Moreover the line bundle $L = \pi^*(K_C \otimes A)$ is nef, since our assumptions guarantee that $\deg(K_C \otimes A) = 2g(C) - 2 + \deg A \geq 0$. If L is nef, then $f^*(L)$ and $f^*(\Omega^1_X)$ are nef (Proposition 3). Since f is a finite surjective map, we finally deduce from Proposition 1 that Ω^1_X is nef.

Remark: (i) The corollary holds for arbitrary compact surfaces too, because Proposition 5 remains true in that case.
(ii) If S is a surface with $\kappa(S) = 1$ and $\pi : S \to \mathbb{P}_1$ is an almost smooth elliptic fibration, the condition that $\deg A \geq 2$ (resp. that L is nef) is automatically satisfied. We have $\deg(\pi_*(\omega_{S/\mathbb{P}_1})) = 0$ and therefore $\pi_*(\omega_{S/\mathbb{P}_1}) = \mathcal{O}_{\mathbb{P}_1}$ (cf. [BPV]). Now the formula for the canonical bundle of an elliptic fibration yields $K_S = \pi^*(K_{\mathbb{P}_1}) \otimes \mathcal{O}_S(D)$, so that $L = K_S$ is nef since $\kappa(S) = 1$.

Similarly we get

COROLLARY 2 *Let X be a projective 3-dimensional manifold with $\kappa(X) \geq 0$ and let $\pi : X \to C$ be an almost smooth fibration, whose fibres are hyperelliptic surfaces. Assume furthermore that* (i) $g(C) \geq 1$ *or* (ii) $g(C) = 0$ *and* $\deg A \geq 2$. *Then Ω^1_X is nef.*

Proof: To deduce from Proposition 6 that $\Omega^1_{X'/C'}$ is nef as a quotient of $\Omega^1_{X'}$, we have to assure that $g(C') \geq 1$. But $g(C') = 0$ leads to $-\infty = \kappa(X') \geq \kappa(X)$ which contradicts our assumptions.

In particular, these two corollaries yield the direction (ii)$\Rightarrow$ (i) in Theorem 3 which is now completely proved.

3 QUOTIENTS WITH NEF COTANGENT BUNDLE

The goal of this section is to prove that compact quotients of a manifold which is hyperbolic with respect to the Carathéodory-Reiffen pseudometric have a nef cotangent bundle. We will use the notion of singular hermitian metrics as introduced in [De1]:

DEFINITION 2 *Let L be a holomorphic line bundle over a compact complex manifold X and let $\theta_\alpha : L|_{U_\alpha} \xrightarrow{\simeq} U_\alpha \times \mathbb{C}$ be a local trivialization of L. Then a singular hermitian metric on L is given by*

$$||\xi|| = |\theta_\alpha(\xi)| \cdot e^{-\varphi_\alpha(x)}, \qquad x \in U_\alpha, \quad \xi \in L_x,$$

where $\varphi_\alpha \in L^1_{loc}(U_\alpha)$ is an arbitrary real valued function, called the weight function of the metric with respect to the trivialization θ_α.

The curvature form of the singular metric on L is locally given by the closed $(1,1)$-current $c(L) = \frac{i}{\pi}\partial\bar{\partial}\varphi_\alpha$. We will write $c(L) \geq 0$, if $c(L)$ is a positive current in the sense of distribution theory, i.e. if the weight functions φ_α are plurisubharmonic.
Remark: We will say that a singular metric is continuous (or simply that it is a continuous metric), if the weight functions φ_α are continuous on the trivialization sets.
The main ingredient for the following arguments will be the next proposition which is independently due to Demailly, Shiffman and Tsuji (see e.g. [De2])

PROPOSITION 7 *Let L be a holomorphic line bundle on a compact complex manifold X. Then L is nef, if there exists a continuous metric with $c(L) \geq 0$.*

In fact the proposition is even true in the case where the Lelong numbers of the metric (which are zero everywhere for a continuous metric) are zero except for a countable set of points (cf. Thm. 4.2 in [JS]).
Let E be a holomorphic vector bundle over a compact complex manifold X. As in [Rei] and [Ko] we define

DEFINITION 3 *A Finsler structure on E is a continuous function $F : E \to \mathbb{R}_{\geq 0}$, so that for all $\eta \in E$:*
(i) *$F(\eta) > 0$ for $\eta \neq 0$,*
(ii) *$F(\lambda\eta) = |\lambda| F(\eta)$ for all $\lambda \in \mathbb{C}$.*
If we require in (i) *only $\geq$, F is said to be a Finsler pseudostructure.*

Let $P(E)$ denote the projective bundle of *lines* in the fibres of E, $p : P(E) \to X$ the projection and $\mathcal{O}_{P(E)}(-1)$ the subbundle of p^*E whose fibre over a point in $P(E)$ is given by the complex line represented by that point. Then we have a map $\tilde{p} : \mathcal{O}_{P(E)}(-1) \to E$ which is biholomorphic outside the zero sections of $\mathcal{O}_{P(E)}(-1)$ and E. The set of all plurisubharmonic functions on a complex manifold Y will be denoted by $PSH(Y)$.

PROPOSITION 8 (a) *Any Finsler structure F on E defines via*

$$||\xi|| := F \circ \tilde{p}(\xi), \qquad \xi \in \mathcal{O}_{P(E)}(-1).$$

a continuous metric on $\mathcal{O}_{P(E)}(-1)$.
(b) *If $\log F \in PSH(E \backslash \{0\})$, then $-\varphi_\alpha \in PSH(U_\alpha)$.*

Proof: (a) Let $\theta_\alpha : \mathcal{O}_{P(E)}(-1) \mid_{U_\alpha} \xrightarrow{\simeq} U_\alpha \times \mathbb{C}$ be a local trivialization and let s_α be a local holomorphic section of $\mathcal{O}_{P(E)}(-1) \mid_{U_\alpha}$ which describes the trivialization. Then the corresponding weight function is

$$-\varphi_\alpha(x) = \log ||s_\alpha(x)|| = \log F \circ \tilde{p}(s_\alpha(x)), \qquad x \in U_\alpha.$$

The map $\tilde{p} \circ s_\alpha : U_\alpha \to E$ is clearly holomorphic. Moreover for $x \in U_\alpha$ we have $s_\alpha(x) \neq 0$, so that property (i) in the definition of Finsler structures leads to $F \circ \tilde{p}(s_\alpha(x)) > 0$. From this we conclude $-\varphi_\alpha \in C^0(U_\alpha)$.
(b) If $f : Y \to Z$ is a holomorphic map between complex manifolds and the function $u \in PSH(Z)$, then $u \circ f \in PSH(Y)$ (cf. [JP], Appendix, PSH 7). So, since $\tilde{p} \circ s_\alpha$ is holomorphic, we have $-\varphi_\alpha \in PSH(U_\alpha)$.

PROPOSITION 9 *Let $E \to X$ be a holomorphic vector bundle over a compact complex manifold X. If there exists a Finsler structure $F : E \to \mathbb{R}_{\geq 0}$ such that* $\log F \in PSH(E\backslash\{0\})$*, then E^* is nef.*

Proof: To prove that E^* is nef, we have to show that $L := \mathcal{O}_{P(E)}(1) \cong \mathcal{O}_{\mathbb{P}(E^*)}(1)$ is nef. According to Proposition 8 the Finsler structure $F : E \to \mathbb{R}_{\geq 0}$ induces a continuous metric on $\mathcal{O}_{P(E)}(-1)$ so that $-\varphi_\alpha \in PSH(U_\alpha)$. For the dual bundle $L = \mathcal{O}_{P(E)}(1)$ equipped with the dual metric the weight functions are given by $\varphi^*_\alpha = -\varphi_\alpha$, hence we have a continuous metric on L whose current is positive and the assertion follows from Proposition 7.

Let X be a connected complex manifold. A Finsler (pseudo-) structure on the tangent bundle TX is called a differential (pseudo-) metric. Any such X admits a differential pseudometric: for $p \in X$ and $\eta \in TX_p$ we define

$$\gamma_X(p, \eta) := \sup\{|Dg(p).\eta| : g \in \mathcal{O}(X, \Delta), g(p) = 0\},$$

where Δ is the open unit disc in $\mathbb{C}$ and $\mathcal{O}(X, \Delta)$ the set of all holomorphic maps from X to Δ. Reiffen shows in [Rei]:

PROPOSITION 10 *The map $\gamma_X : TX \to \mathbb{R}_{\geq 0}$ is a differential pseudometric, which has the following invariance property. Let $f : X \to Y$ be a holomorphic map of connected complex manifolds, then*

$$\gamma_Y(f(p), Df(p).\eta) \leq \gamma_X(p, \eta),$$

in particular, for a biholomorphic map f the equality holds.

The function γ_X is called the Carathéodory-Reiffen pseudometric and X is said to be γ-hyperbolic, if γ_X is a differential metric.
Examples: (i) Any bounded domain $G \subset \mathbb{C}^n$ is γ-hyperbolic (cf. [JP], Chap. II, Prop. 2.3.2).
Proposition 10 immediately implies: let $i : X \to Y$ be a holomorphic immersion and let Y be γ-hyperbolic, then X is γ-hyperbolic too. This gives us
(ii) Let Y be a Stein manifold and let $\tilde{G}$ be a bounded domain in Y, i.e. there exists an embedding $Y \hookrightarrow \mathbb{C}^N$ and a bounded domain $G \subset \mathbb{C}^N$, such that $\tilde{G} = Y \cap G$ is connected. Then $\tilde{G}$ is γ-hyperbolic.

PROPOSITION 11 *Let X be a γ-hyperbolic manifold. Then the function*

$$\log \gamma_X : TX\backslash\{0\} \to (-\infty, +\infty)$$

is plurisubharmonic.

Proof: Since the logarithm is strictly increasing, we have

$$\log \gamma_X(p, \eta) = \sup\{\log |Dg(p).\eta| : g \in \mathcal{O}(X, \Delta), g(p) = 0\}.$$

The tangent map of a holomorphic map is again holomorphic, so that $\tilde{g}(p, \eta) := \log |Dg(p).\eta|$ is in $PSH(TX)$ (see [JP], Appendix, PSH 4). Hence $\log \gamma_X = \sup_g\{\tilde{g}\}$ is the supremum of plurisubharmonic functions. By assumption γ_X is a differential metric, i.e. γ_X is continuous and $\gamma_X : TX\setminus\{0\} \to \mathbb{R}_{>0}$, thus $\log \gamma_X : TX\setminus\{0\} \to (-\infty, \infty)$ is also continuous. Now we get our assertion from the following fact ([JP], Appendix, PSH 14). If a family $(u_\alpha)_{\alpha\in A}$ of plurisubharmonic functions is locally uniformly bounded from above, then the function

$$u_0 := (\sup_{\alpha\in A} u_\alpha)^*$$

is again plurisubharmonic, where "$*$" denotes the upper semicontinuous regularization. But we don't need to regularize $\log \gamma_X$, since it is already continuous and this assures also that the family $\{\tilde{g}\}$ is locally uniformly bounded from above.
Let $\mathcal{G}$ be a subgroup of the automorphism group $\mathrm{Aut}(X)$, which operates fixpointfree and properly discontinuously on X. Then the quotient $Q = X/\mathcal{G}$ is a Hausdorff space which admits a unique complex structure, such that the projection $\pi : X \to Q$ is a holomorphic and locally biholomorphic map. We can now prove

THEOREM 6 *Let X be a γ-hyperbolic manifold and let $Q = X/\mathcal{G}$ be a compact quotient as above. Then the cotangent bundle Ω^1_Q is nef.*

Proof: As local coordinates ψ for Q we can take π^{-1} restricted to appropriate open sets such that a coordinate change is described by $\psi_1 \circ \psi_0^{-1} = f$, where $f \in \mathcal{G}$ (cf. [W], Chap. V, Prop. 5.3.). Then we define for $q \in Q$ and $\xi \in TQ_q$

$$F(q,\xi) := \gamma_X(\psi(q), D\psi(q).\xi).$$

Since the Carathéodory-Reiffen metric γ_X is invariant under biholomorphic transformations (Proposition 10), this definition does not depend on the choice of the local coordinate and gives us a differential metric F on TQ. Moreover Proposition 11 implies that $\log F \in PSH(TQ\setminus\{0\})$. Now the assertion follows from Proposition 9.

In particular, compact quotients of a bounded domain in $\mathbb{C}^n$ or in a Stein manifold have nef cotangent bundles.

REFERENCES

[BPV] Barth, W., Peters, C., Van de Ven, A.: Compact Complex Surfaces. Erg. d. Math. Bd. 3, Berlin Heidelberg New York: Springer 1984

[Ca] Campana, F.: Réduction d'Albanese d'un morphisme propre et faiblement kählérien. I. Comp. Math. 54, 373-398 (1985)

[CP] Campana, F., Peternell, T.: Projective manifolds whose tangent bundles are numerically effective. Math. Ann. 289, 169-187 (1991)

[De1] Demailly, J.-P.: Singular hermitian metrics on positive line bundles. In: Hulek, K. et al. (eds.): Complex Algebraic Varieties (Lect. Notes Math., vol. 1507, pp. 87-104) Berlin Heidelberg New York: Springer 1992

[De2] Demailly, J.-P.: Regularization of closed positive currents and intersection theory. J. Alg. Geom. 1, 361-409 (1992)

[DPS] Demailly, J.-P., Peternell, T., Schneider, M.: Compact complex manifolds with numerically effective tangent bundles. J. Alg. Geom. 3, 295-345 (1994)

[Fu] Fujiwara, T.: Varieties of small Kodaira dimension whose cotangent bundles are semiample. Comp. Math. 84, No.1, 43-52 (1992)

[Gr] Griffiths, P.: Periods of integrals on algebraic manifolds, III. Publ. Math. I.H.E.S. No. 38, 125-180 (1970)

[Ha] Hartshorne, R.: Algebraic Geometry. Graduate Texts in Math. 52, Berlin Heidelberg New York: Springer 1977

[JP] Jarnicki, M., Pflug, P.: Invariant Distances and Metrics in Complex Analysis. De Gruyter Expositions in Math. 9, Berlin New York: De Gruyter 1993

[JS] Ji, K., Shiffman, B.: Properties of compact complex manifolds carrying closed positive currents. J. Geom. Anal. (1993)

[Ko] Kobayashi, S.: Negative vector bundles and complex Finsler structures. Nagoya Math. J., Vol. 57, 153-166 (1975)

[Kod] Kodaira, K.: On Compact Analytic Surfaces: II. Ann. Math. 77, 563-626 (1963)

[Mo] Mori, S.: Projective manifolds with ample tangent bundles. Ann. Math. (2) 110, 593-606 (1979)

[PW] Peternell, T., Wilson, P.M.H.: Threefolds with extremal Chern classes. Higher dimensional complex varieties, Trento Proceedings, 357-378 (1996)

[Re] Reid, M.: Bogomolov's theorem $c_1^2 \leq 4c_2$. Intl. Symp. on Algebraic Geometry Kyoto, 623-642 (1977)

[Rei] Reiffen, H.J.: Die Carathéodorysche Distanz und ihre zugehörige Differentialmetrik. Math. Ann. 161, 315-324 (1965)

[Ue] Ueno, K.: Classification Theory of Algebraic Varieties and Compact Complex Spaces. Lect. Notes Math., vol. 439, Berlin Heidelberg New York: Springer 1975

[W] Wells, R.O.: Differential Analysis on Complex Manifolds. Graduate Texts in Math. 65, Berlin Heidelberg New York: Springer 1980

[Zh] Zheng, F.: On semi-positive threefolds. Thesis, Harvard (1990)

Henrik Kratz
Mathematisches Institut
Universität Bayreuth
95440 Bayreuth
Germany
kratz@btm8x1.mat.uni-bayreuth.de

Maps onto Certain Fano Threefolds

Ekaterina Amerik

Received: May 15, 1997
Revised: July 11, 1997

Communicated by Thomas Peternell

Abstract. We prove that if X is a smooth projective threefold with $b_2 = 1$ and Y is a Fano threefold with $b_2 = 1$, then for a non-constant map $f : X \to Y$, the degree of f is bounded in terms of the discrete invariants of X and Y. Also, we obtain some stronger restrictions on maps between certain Fano threefolds.

1991 Mathematics Subject Classification: 14E99, 14J45

1. Introduction

Let X, Y be smooth complex n-dimensional projective varieties with $Pic(X) \cong Pic(Y) \cong \mathbf{Z}$. Let $f : X \to Y$ be a non-constant morphism. It is a trivial consequence of Hurwitz's formula

$$K_X = f^*K_Y + R$$

that if Y is a variety of general type, then $deg(f)$ is bounded in terms of the numerical invariants of X and Y, and in particular all the morphisms from X to Y fit in a finite number of families.

If we drop the assumption that Y is of general type, then this assertion is no longer quite true. Indeed, if Y is a projective space $\mathbf{P}^n$, for any X we can construct infinitely many families of maps $X \to Y$: take an embedding of X in $\mathbf{P}^N$ by any very ample divisor on X and then project the image to $\mathbf{P}^n$. However, one might ask if $\mathbf{P}^n$ is the only variety with this property (the following conjectures are suggested by A. Van de Ven) :

Conjecture A: *Let X, Y be as above and $Y \ncong \mathbf{P}^n$. Then there is only finitely many families of maps from X to Y. Moreover, the degree of a map $f : X \to Y$ can be bounded in terms of the discrete invariants of X and Y.*

A weaker version is the following

Conjecture B: *Let X, Y be smooth n-dimensional projective varieties with $b_2(X) = b_2(Y) = 1$. Suppose $Y \not\cong \mathbf{P}^n$ and, if $n = 1$, that Y is not an elliptic curve. Then the degree of a map $f : X \to Y$ can be bounded in terms of the discrete invariants of X and Y.*

Remark: If $n = 1$, the Conjecture A is empty and the Conjecture B is trivial. If $n = 2$, one must check the Conjecture A with Y a K3-surface, and at the moment I do not know how to do this. This problem, of course, does not arise for Conjecture B, which again becomes a triviality in dimension two (note that if for a smooth complex projective variety V we have $b_1(V) \neq 0$ and $b_2(V) = 1$, then V is a curve). The assumption in the Conjecture B that Y is not an elliptic curve is , of course, necessary: any torus has endomorphisms of arbitrarily high degree given by multiplication by an integer.

Evidence: It seems likely that "the more ample is the canonical sheaf on Y, the more difficult it becomes to produce maps from X to Y". Of course, the projective space has the "least ample" canonical sheaf: $K_{\mathbf{P}^n} = -(n+1)H$, where H is a hyperplane. The next case is that of a quadric: $K_{Q_n} = -nH$ with H a hyperplane section. For $n = 3$, it has been proved by C.Schuhmann ([S]) that the degree of a map from a smooth threefold X with Picard group $\mathbf{Z}$ to the three-dimensional quadric is bounded in terms of the invariants of X. In [A], I have suggested a simpler method to prove results of this kind, which also generalizes to higher dimensions.

The main purpose of this paper is to show by a rather simple method that for Fano threefolds Y, at least for those with very ample generator of the Picard group, the above Conjecture B is true (we also show that for many of such threefolds Conjecture A holds). The boundedness results are proved in the next section. In Section 3, we obtain in a similar way a strong restriction on maps between "almost all" Fano threefolds with Picard group $\mathbf{Z}$. This is related to the "index conjecture" of Peternell which states that if $f : X \to Y$ is a map between Fano varieties of the same dimension with cyclic Picard group, then the index of Y is not smaller than that of X. This conjecture is studied for Fano threefolds by C.Schuhmann in her thesis, and one of her main results is that there are no maps from such a Fano threefold of index two to a Fano threefold of index one with reduced Hilbert scheme of lines. An extension of this result appears also in Theorem 3.1 of this paper ; however, there is at least one Fano threefold of index one with non-reduced Hilbert scheme of lines, namely, Mukai and Umemura's V_{22}. The last section of this paper deals with this variety: it is proved that a Fano threefold of index two with Picard group $\mathbf{Z}$ does not admit a map onto it. One would think that the Mukai-Umemura V_{22} is the only Fano threefold of genus at least four with cyclic Picard group and non-reduced Hilbert scheme of lines. The proof of this would be a solution to the "index conjecture" in the three-dimensional case (recall that a Fano threefold of index one and genus at most three has the third Betti number which is bigger than the third Betti number of any Fano threefold of index two ([I1] ,table 3.5), so we do not have to consider the case of genus less than four to prove the index conjecture). In fact even a weaker statement would suffice (see Theorem 3.1).

This paper can be viewed as a very extensive appendix to [A], as a large part of the method is described there.

We will often use the following notations: Generally, for $X \subset \mathbf{P}^n$, H_X denotes the hyperplane section divisor on X. Also, for X with cyclic Picard group, we will call H_X the ample generator of $Pic(X)$ (in this paper it will mostly be assumed that H_X is very ample). By V_k, following Iskovskih, we will often denote a Fano threefold with cyclic Picard group, which has index one and for which $H_X^3 = k$ (k will be called the degree of this Fano threefold). For Grassmann varieties, we use projective notation: $G(k, n)$ denotes the variety of projective k-subspaces in the projective n-space.
Finally, throughout the paper we work over the field of complex numbers.

ACKNOWLEDGMENTS: I would like to thank Professor A. Van de Ven for many helpful discussions. I am grateful to Frank-Olaf Schreyer for explaining me many facts on V_{22} and for letting me use his unfinished manuscript [Sch], and also to Aleksandr Kuznetsov for giving me his master's thesis [K]. The final version of this paper was written during my stay at the University of Bayreuth, to which I am grateful for its hospitality and support.

2. BOUNDEDNESS

Let Y be a Fano threefold such that $Pic(Y) \cong \mathbf{Z}$, and suppose that the positive generator of the Picard group is very ample. When speaking of $deg(Y)$ and other notions related to the projective embedding (e.g. the sectional genus $g(Y)$ of Y) we will suppose that this embedding is given by global sections of the generator.
It is well-known ([I],I, section 5) that if Y is of index two, then lines on Y are parameterized by a smooth surface F_Y (the Fano surface) on Y. A general line on Y has trivial normal bundle, and there is a curve on F which parametrizes lines with the normal bundle $\mathcal{O}_{\mathbf{P}^1}(-1) \oplus \mathcal{O}_{\mathbf{P}^1}(1)$ (let us call them (-1,1)-lines). If Y is of index one, than Y contains a one-dimensional family of lines ([I], II, section 3); the normal bundle of a line is then either $\mathcal{O}_{\mathbf{P}^1}(-1) \oplus \mathcal{O}_{\mathbf{P}^1}$, or $\mathcal{O}_{\mathbf{P}^1}(-2) \oplus \mathcal{O}_{\mathbf{P}^1}(1)$. In the last case such a line is of course a singular point of the Hilbert scheme. In the sequel we will use the simple fact that if the Hilbert scheme of lines on a Fano threefold of index one is non-reduced, i.e. every line of one of its irreducible components is (-2,1), then the surface covered by the lines of this component is either a cone, or a tangent surface to a curve.
If the generator H_Y of $Pic(Y)$ is not very ample, there still exist "lines" on Y: we call a curve C a line if $C \cdot H_Y = 1$. In this case, however, there exist other possibilities for the normal sheaf $N_{C,Y}$. If Y is a threefold of index 2 and $H_Y^3 = 1$, C can even be a singular curve and, moreover, if we want our "lines" to fit into a Hilbert scheme, we must also allow embedded points ([T]).
At this point, it is convenient to recall from [I] which Fano threefolds have very ample/not very ample generator of the Picard group. For index two, the threefolds with very ample generator are cubics, intersections of two quadrics and the linear section of $G(1, 4)$; the other threefolds are double covers of $\mathbf{P}^3$ branched in a quartic (quartic double solids) and double covers of the Veronese cone branched in a cubic section of it (double Veronese cones). For index one, we have nine families of threefolds

with very ample generators, plus double covers of the quadric branched in a quartic section and double covers of $\mathbf{P}^3$ branched in a sextic.
Often we will assume here for simplicity that H_Y is very ample, and discuss the other case in remarks.
We start by proving the following

Proposition 2.1 *A) If Y is a Fano threefold (with $Pic(Y) \cong \mathbf{Z}$, H_Y very ample) of index 2 such that the surface $U_Y \subset Y$ which is the union of all (-1,1)-lines on Y is in the linear system $|iH_Y|$ with $i \geq 5$, then for any threefold X, $Pic(X) \cong \mathbf{Z}$, the degree of a map $f : X \to Y$ is bounded in terms of the discrete invariants of X.*
B) If Y is a Fano threefold of index 1 with $Pic(Y) \cong \mathbf{Z}$, H_Y very ample, such that the surface $S_Y \subset Y$ which is the union of all lines on Y is in the linear system iH_Y with $i \geq 3$, then for any threefold X, $Pic(X) \cong \mathbf{Z}$, the degree of a map $f : X \to Y$ is bounded in terms of the discrete invariants of X.

Proof: Let m be such that $f^*H_Y = mH_X$. Notice that by Hurwitz' formula, our conditions on U_Y resp. S_Y just mean that if $deg(f)$ is big enough, then not the whole inverse image of U_Y resp. S_Y is contained in the ramification. Indeed, if Y is, say, of index one, we have $K_Y = -H_Y$. The Hurwitz formula reads

$$K_X = -mH_X + R.$$

If the whole inverse image of S_Y is in the ramification, then R is at least $\frac{3}{2}mH_X$, so m cannot get very big. Therefore one gets that the inverse image D of a general (-1,1)-line on Y (in the index-two case) or a general line on Y (in the index-one case) has a reduced irreducible component C.
Let Y be a Fano threefold of index two satisfying $U_Y = iH_Y$ with $i \geq 5$. For C and D as above, there is a natural morphism

$$\phi : (\mathcal{I}_C/\mathcal{I}_C^2)^* \to (\mathcal{I}_D/\mathcal{I}_D^2)^*|_C = \mathcal{O}_C(m) \oplus \mathcal{O}_C(-m),$$

and this map must be an isomorphism at a smooth point of D, i.e. at a sufficiently general point of C, as C is reduced. Now, also due to the fact that C is reduced, the natural map

$$\psi : T_X|_C \to (\mathcal{I}_C/\mathcal{I}_C^2)^*$$

is a generic surjection. Therefore if we find an integer j such that $T_X(j)$ is globally generated, we must have $m \leq j$.
Such j depends only on the discrete invariants of X. Indeed, let A be a very ample multiple of H_X. A linear subsystem of the sections of A gives an embedding of a threefold X into $\mathbf{P}^7$. We have

$$T_X(K_X) = \Lambda^2\Omega_X.$$

$\Lambda^2\Omega_X$ is a quotient of $\Lambda^2\Omega_{\mathbf{P}^7}|_X$, and we deduce from this that $\Lambda^2\Omega_X(3A)$ is generated by the global sections. So $T_X(K_X + 3A)$ is generated by the global sections, and j can be taken such that $K_X + 3A = jH_X$. So one only needs to know which multiple of H_X is very ample, and this can be expressed in terms of the discrete invariants of X (see for example [D] for many results in this direction).

The case of index one is completely analogous: a normal bundle of any line on a Fano threefold of index one has a negative summand.

REMARK A: The assumption on the very ampleness of the generator of $Pic(Y)$ is not really necessary to prove Proposition 2.1. Otherwise, we call "lines" curves C satisfying $C \cdot H_Y$=1. These curves are rational. One has then to count with the possibility that e. g. some of the "lines" on such a Fano 3-fold of index two can have normal bundle $\mathcal{O}_{\mathbf{P}^1}(-2) \oplus \mathcal{O}_{\mathbf{P}^1}(2)$, but this is not really essential for the argument: as soon as we can find sufficiently big 1-parameter family of smooth rational curves with a negative summand in the normal bundle, our method works.

EXAMPLES OF FANO THREEFOLDS Y SATISFYING OUR ASSUMPTIONS ON S_Y, U_Y:

1) Y a cubic in $\mathbf{P}^4$ and
2) Y an intersection of two quadrics in $\mathbf{P}^5$. To check this is more or less standard and almost all details can be found in [CG] for a cubic and in [GH] (Chapter 6) for an intersection of two quadrics. For convenience of the reader, we give here the argument for Y an intersection of two quadrics in $\mathbf{P}^5$:
Let $F \subset G(1,5)$ be a surface which parametrizes lines on Y (Fano surface) , and let $\mathcal{U} \to F$ be the family of these lines. The ramification locus of the natural finite map $\mathcal{U} \to Y$ consists exactly of (-1,1)-lines, that is, the surface M covered by (-1,1)-lines on Y is exactly the set of points of Y through which there pass less than four lines (of course there are four lines through a general point of Y). F is the zero-scheme of a section of the bundle $S^2U^* \oplus S^2U^*$ on $G(1,5)$. A standard computation with Chern classes yields then that $K_F = \mathcal{O}_F$ (in fact, F is an abelian variety ([GH])). For a general line $l \subset Y$ consider a curve $C_l \subset F$ which is the closure in F of lines intersecting l and different from l. C_l contains l iff l is (-1,1). C_l is smooth for any l ([GH]). By adjunction, C_l has genus 2. So the ramification R of the natural 3:1 morphism $h_l : C_l \to l$ sending l' to $l \cap l'$ (with l general, i.e. not a (-1,1)-line) has degree 8. The branch locus of h consists of intersection points of l and the surface M of (-1,1)-lines, and so we have that this surface is in $|iH_Y|$ with $i \geq 4$ and $i = 4$ only if there are only 2 lines through a general point of M. This is again impossible: otherwise, for l a (-1,1)-line, C_l would be birational to l. In fact, one gets that $i = 8$.
3) Y a quartic double solid. The computations are rather similar, and the best reference is [W]. Bitangent lines to the quartic surface give pairs of "lines" on Y as their inverse images under the covering map. Welters proves the following results: the Fano surface F_Y has only isolated singularities (and is smooth for a general Y); the curve C_l for a general l is smooth except for one double point; there are 12 "lines" through a general point of Y; $p_a(C_l) = 71$. We use these results to conclude that Y satisfies our assumptions.
4) Y is a "sufficiently general" Fano threefold of index one (of course we assume that $Pic(Y) \cong \mathbf{Z}$ and that the positive generator of $Pic(Y)$ is very ample), $deg(Y) \neq 22$: see [I], II, proof of th. 6.1. It is computed there that a Fano threefold Y of index one (with very ample H_Y) with reduced scheme of lines satisfies our assumption on S_Y iff $deg(Y) \neq 22$. By the classification of Mukai ([M]), any Fano threefold of index one as above except V_{22}'s is a hyperplane section of a smooth (Fano) fourfold. Clearly, a general line on a Fano fourfold of index two has trivial normal bundle. So a general

hyperplane section of such a fourfold has reduced Hilbert scheme of lines.
5) Y any Fano threefold of index one and genus 10: Prokhorov shows in [P] that the Hilbert scheme of lines on *any* such threefold is reduced.
6) Y any Fano threefold V_{14} of index one and genus 8: such a threefold is a linear section of $G(1,5)$ in the Plücker embedding. Iskovskih shows in [I], II, proof of th. 6.1 (vi), that on such a threefold with reduced scheme of lines, lines will cover a surface which is linearly equivalent to $5H$. So one sees that if the lines cover only H or $2H$, the scheme of lines is non-reduced and the surface covered by lines consists of one or two components which are hyperplane sections of Y. Moreover, as a V_{14} does not contain cones, all the lines in one of the components must be tangent to some curve A. One checks easily that this curve is a rational normal octic. A is then the Gauss image of a rational normal quintic B in $\mathbf{P}^5$ ([A], proof of Proposition 3.1(ii)). This makes it possible to check that there is no smooth three-dimensional linear section of $G(1,5)$ containing the tangent surface to A. Indeed, one can assume that B is given as

$$(x_0^5 : x_0^4 x_1 : \ldots : x_1^5), (x_0 : x_1) \in \mathbf{P}^1;$$

one computes then that the Gauss image of B in $G(1,5) \subset \mathbf{P}^{14}$ (where $G(1,5)$ is embedded to $\mathbf{P}^{14}$ by Plücker coordinates (z_i), the order of which we take as follows: for a line l through $p = (p_0 : \ldots : p_5)$ and $q = (q_0 : \ldots : q_5)$ we take $z_0 = p_0q_1 - p_1q_0$; $z_1 = p_0q_2 - p_2q_0$; ...; $z_5 = p_1q_2 - p_2q_1$; ...; $z_{14} = p_4q_5 - p_5q_4$) generates the linear subspace L given by the following equations:

$$z_2 = 3z_5, z_3 = 2z_6, z_4 = 5z_9,$$
$$z_7 = 3z_9, z_8 = 2z_{10}, z_{11} = 3z_{12}.$$

So we must consider all the projective 9-subspaces through L and prove that the intersection of every such space with $G(1,5)$ is singular. This can be done for example as follows: let $\mathcal{L} \cong \mathbf{P}^5$ be a parametrizing variety for these 9-subspaces. Notice that the points $x = (1 : 0 : \ldots : 0)$ and $y = (0 : \ldots : 0 : 1)$ belong to our curve A. Notice that if t is a point of A, then the set $\mathcal{L}_t = \{M \in \mathcal{L} : M \cap G(1,5) \text{ is singular at } t\}$ is a hyperplane in $\mathcal{L}$. If we see that these sets are different at different points t, we are done. It is not difficult to check explicitly (writing down the matrix of partial derivatives) that for $x = (1 : 0 : \ldots : 0) \in A$ and $y = (0 : \ldots : 0 : 1) \in A$, $\mathcal{L}_x \neq \mathcal{L}_y$: if a 9-space M through L is given by the equations

$$a_{1i}(z_2 - 3z_5) + a_{2i}(z_3 - 2z_6) + a_{3i}(z_7 - 3z_9) +$$
$$+ a_{4i}(z_8 - 2z_{10}) + a_{5i}(z_{11} - 3z_{12}) + a_{6i}(z_4 - 5z_9) = 0$$

for $i = 1, \ldots, 5$, then $M \in \mathcal{L}_x$ if and only if

$$det(a_{ki})_{k=1,2,3,4,6}^{i=1,2,3,4,5} = 0$$

and $M \in \mathcal{L}_y$ if and only if

$$det(a_{ki})_{k=1,2,3,4,5}^{i=1,2,3,4,5} = 0.$$

These conditions are clearly different.

EXAMPLES OF FANO THREEFOLDS NOT SATISFYING ASSUMPTIONS OF PROPOSITION 2.1:

1) Y is a linear section of $G(1,4)$ in the Plücker embedding: the surface U_Y has degree 10.
2) Y is a Fano variety of index one and genus 12 (V_{22}). The surface of lines belongs to $|-2K_Y|$ for all V_{22}'s but one ([P]), for which the scheme of lines is non-reduced and the surface covered by lines belongs to $|-K_Y|$. This threefold with non-reduced Hilbert scheme of lines (the Mukai-Umemura variety) will be denoted V_{22}^s.

QUESTION: *Are these the only examples?*

REMARK B: Though any V_{22} violates the assumption of the Proposition 2.1, for a V_{22} with the reduced Hilbert scheme of lines (therefore for all V_{22}'s but one) the boundedness of the degree of a map $f: X \to V_{22}$ can be proved. The point is that a general line on such a V_{22} has the normal bundle $\mathcal{O}_{\mathbf{P}^1} \oplus \mathcal{O}_{\mathbf{P}^1}(-1)$, so if U is the universal family of lines on V_{22} and $\pi: U \to V_{22}$ is the natural map, then π is an immersion along a general line. Now if the preimage of a general line l is not contained in the ramification R, one can proceed as before. If it is, then let C be the reduction of an irreducible component of $f^{-1}(l)$, and let k be such that at a general point of the component of R containing C, the ramification index is $k-1$ (i.e. "k points come together".) It turns out that using our observation about π, we can then estimate the arithmetic genus of C (see [A], section 5). Namely, let $f^*H_{V_{22}} = mH_X$ and let $K_X = rH_X$. We get then

$$2p_a(C) - 2 \leq (r - \frac{m}{k})CH_X.$$

Suppose now that $k-1$ is a smallest ramification index for R. Hurwitz' formula implies that if $r < \frac{m}{3}$, then $k = 2$. So if m gets big, $p_a(C)$ becomes negative, and this is impossible.

Concerning the remaining Fano threefolds (in particular, V_{22}^s and $G(1,4) \bigcap \mathbf{P}^6$), we can prove a weaker result (as in Conjecture B):

PROPOSITION 2.2 *Let Y be a Fano threefold with $Pic(Y) = \mathbf{Z}$ and with H_Y very ample, let X be a smooth threefold with $b_2(X) = 1$ and let $f: X \to Y$ be a morphism. If either Y is of index two, or Y is of index one with non-reduced Hilbert scheme of lines, then the degree of f is bounded in terms of the discrete invariants of X.*

Proof: Consider for example the index one case. We have that Y has a one-dimensional family of $(-2,1)$-lines. If we take a smooth hyperplane section H through a line l of this family, the sequence of the normal bundles

$$0 \to N_{l,H} \to N_{l,Y} \to N_{H,Y}|_l \to 0$$

splits.
Therefore, if M is the inverse image of H and C is the inverse image of l (scheme-theoretically), the sequence

$$0 \to N_{C,M} \to N_{C,X} \to N_{M,X}|_C \to 0$$

also splits.
It is not difficult to see that for a general choice of l and H, the surface M has only isolated singularities. As M is a Cartier divisor on a smooth variety X (say $M \in |\mathcal{O}_X(m)|$), M is normal.
Now we are in the situation which is very similar to that of the following

Theorem (R. Braun, [B]): *Let W be a Cartier divisor on a variety V of dimension n, $2 \le n < N$, in $\mathbf{P}^N$ such that W has an open neighborhood in V which is locally a complete intersection in $\mathbf{P}^N$. If the sequence of the normal bundles*

$$0 \to N_{W,V} \to N_{W,\mathbf{P}^N} \to N_{V,\mathbf{P}^N}|_W \to 0 \ (*)$$

splits, then W is numerically equivalent to a multiple of a hyperplane section of V.

It turns out that if we replace here W, V, $\mathbf{P}^N$ by C, M, X as in our situation, the similar statement is true. The only additional assumption we must make is that M is sufficiently ample, i.e. m is sufficiently big:
Claim: Let X be a smooth projective 3-fold with $b_2(X) = 1$, and let M be a sufficiently ample normal Cartier divisor on X. If C is a Cartier divisor on M and the sequence

$$0 \to N_{C,M} \to N_{C,X} \to N_{M,X}|_C \to 0$$

splits, then C is numerically equivalent to a multiple of $H_X|_M$.
The proof of this claim is almost identical to that of Braun's theorem (which is itself a refinement of the argument of [EGPS] where the theorem is proved for V a smooth surface). Recall that the main steps of this proof are:

1) The sequence $(*)$ splits iff W is a restriction of a Cartier divisor from the second infinitesimal neighborhood V_2 of V in $\mathbf{P}^N$;
2)The image of the natural map $Pic(V_2) \to Num(V)$ is one-dimensional.

In the situation of the lemma, 1) goes through without changes with W, V, $\mathbf{P}^N$ replaced by C, M, X (M_2 will of course denote the second infinitesimal neighborhood of M in X). The second step is an obvious modification of that in [B], [EGPS]: as in these works, it is enough to prove that the image of the natural map

$$Pic(M_2) \to H^1(M, \Omega^1_M)$$

is contained in a one-dimensional complex subspace, and this follows from the commutative diagram

$$\begin{tikzcd}
Pic(M_2) \arrow[r, "restr."] \arrow[d, "dlog"] & Pic(M) \arrow[r] & Num(M) \arrow[r] & H^1(M, \Omega^1_M) \\
H^1(M_2, \Omega^1_{M_2}) \arrow[r] & H^1(M, \Omega^1_{M_2}|_M) \arrow[r, "\alpha"] & H^1(M, \Omega^1_X|_M) \arrow[ur] &
\end{tikzcd}$$

(where α exists because the sheaves $\Omega^1_{M_2}|_M$ and $\Omega^1_X|_M$ are isomorphic) and the fact that for sufficiently ample M,

$$H^1(M, \Omega^1_X|_M) \cong H^1(X, \Omega^1_X) \cong \mathbf{C}$$

as follows from the restriction exact sequence

$$0 \to \Omega^1_X(-M) \to \Omega^1_X \to \Omega^1_X|_M \to 0.$$

Note that we can give an effective estimate for "sufficient ampleness" of M in terms of numerical invariants of X using the Griffiths vanishing theorem ([G]).
Applying this to our situation of a map onto a Fano threefold Y of index one with non-reduced Hilbert scheme of lines, we get that $C = f^{-1}(l)$ must be numerically equivalent to a multiple of the hyperplane section divisor on $M = f^{-1}(H)$ if the number m (defined by $f^*H_Y = mH_X$) is large enough. As it is easy to show that C and the hyperplane section of M are independent in $Num(M)$, it follows that m and therefore $deg(f)$ must be bounded. The case of index two is exactly the same (use the existence of a divisor covered by (-1,1)-lines). So the Proposition is proved.

We summarize our results in the following

Theorem 2.3 *Let X be a smooth projective threefold with $b_2(X) = 1$, let Y be a Fano threefold with $b_2(Y) = 1$ and very ample H_Y and let $f : X \to Y$ be a morphism. If $Y \not\cong \mathbf{P}^3$, then the degree of f is bounded in terms of the discrete invariants of X, Y.*

Proof: Indeed, there are only four possibilities:
a) Y is a quadric: this is proved in [S], [A].
b) Proposition 2.1 applies;
c) Y is V_{22} with reduced scheme of lines: the boundedness for $deg(f)$ is obtained in Remark B;
d) Y is either $G(1,4) \cap \mathbf{P}^6$, or a Fano threefold with non-reduced Hilbert scheme of lines: then Proposition 2.2 applies.
Notice that in the first three cases it is sufficient that $Pic(X) \cong \mathbf{Z}$.

3. Maps between Fano threefolds

It turns out that we obtain especially strong bound if X is also a Fano variety. In many cases,this even implies non-existence of maps:

Theorem 3.1 *Let X, Y be Fano threefolds, $Pic(X) \cong Pic(Y) \cong \mathbf{Z}$. Suppose that H_X, H_Y are very ample. If either*
i) Y is of index one and S_Y is at least $2H_Y$,
or
ii) Y is of index two and U_Y is at least $4H_Y$(where S_Y, U_Y are as in Proposition 2.1), then for a non-constant map $f : X \to Y$ we must have

$$f^*(H_Y) = H_X,$$

i.e.

$$deg(f) = \frac{H_X^3}{H_Y^3}.$$

Before starting the proof, we formulate the following result from [S]:
Let $f : X \to Y$ be a non-trivial map between Fano threefolds with Picard group $\mathbf{Z}$*. Then:*
A) If X, Y are of index two, then the inverse image of any line is a union of lines;
B) If X, Y are of index one, then the inverse image of any conic is a union of conics;
C) If X is of index one and Y is of index two, then the inverse image of any line is a union of conics;
D) If X is of index two and Y is of index one, then the inverse image of any conic is a union of lines.
(here a conic is allowed to be reducible or non-reduced. Unions of lines and conics are understood in set-theoretical sense, i.e. a line or a conic from this union can, of course, have a multiple structure.)

We will also need some facts on conics on a Fano threefold V of index one, with very ample $-K_V$ and cyclic Picard group. Iskovskih proves ([I],II, Lemma 4.2) that if C is a smooth conic on such a threefold, then $N_{C,V} = \mathcal{O}_{\mathbf{P}^1}(-a) \oplus \mathcal{O}_{\mathbf{P}^1}(a)$ with a equal to 0,1,2 or 4. The following lemma is an almost obvious refinement of this:

Lemma 3.2 *a) Let $C \subset V$ be a smooth conic. Then $N_{C,V} = \mathcal{O}_{\mathbf{P}^1}(-4) \oplus \mathcal{O}_{\mathbf{P}^1}(4)$ if and only if there is a plane tangent to V along C. In particular, such conics exist only if V is a quartic.*
b) Let $C \subset V$ be a reducible conic: $C = l_1 \bigcup l_2$, $l_1 \neq l_2$. Let N be the (locally free with trivial determinant) normal sheaf of C in V. Then $N|_{l_i} = \mathcal{O}_{\mathbf{P}^1}(-a_i) \oplus \mathcal{O}_{\mathbf{P}^1}(a_i)$ with $0 \leq a_i \leq 2$, and if $a_i = 2$ for both i, then there is a plane tangent to V along C (and V is a quartic).

Proof: a) This is a simple consequence of the fact that for $C \subset V \subset \mathbf{P}^n$, $N_{C,V} \subset N_{C,\mathbf{P}^n}$, and the only subbundle of degree 4 in $N_{C,\mathbf{P}^n}$ is $N_{C,P}$ with P the plane containing C. One concludes that V is a quartic as all the other Fano threefolds V considered here are intersections of quadrics and cubics which contain this V ([I], II, sections 1,2) and therefore must contain this P, which is impossible.
b) We have embeddings

$$0 \to N_{l_i,V} \to N|_{l_i},$$

this implies the first statement: $0 \leq a_i \leq 2$. If $a_i = 2$, then l_i should be a (-2,1)-line; therefore there are planes P_i tangent to V along l_i, giving the degree 1 subbundle of $N_{l_i,V}$ and the exceptional section in $\mathbf{P}(N_{l_i,V}) \cong F_3$. In fact $P_1 = P_2$. This is easy to see using so-called " elementary modifications" of Maruyama (of which I learned from [AW] ,p.11): if we blow $\mathbf{P}(N_{l_1,V})$ up in the point p corresponding to the direction of l_2 and then contract the proper preimage of the fiber, we will get $\mathbf{P}(N|_{l_1})$. Under our circumstances, p must lie on the exceptional section of $\mathbf{P}(N_{l_1,V})$, so $l_2 \subset P_1$. In the same way, $l_1 \subset P_2$, q.e.d..

Proof of the Theorem:
Let $f : X \to Y$ be a finite map between Fano threefolds as above.
Again, the condition on S_Y, T_Y means that not the whole inverse image of S_Y, T_Y can be contained in the ramification. If Y is of index one resp. index two, we will denote by C be a reduced irreducible component of the inverse image of a general line

resp. (-1,1)-line l on Y (so C is not contained in the ramification), and by D the full scheme-theoretic inverse image of such a line.
Let $f^*\mathcal{O}_Y(1) = \mathcal{O}_X(m)$. If X is of index two, then $T_X(1)$ is globally generated. As in the Proposition 2.1, we conclude that $m = 1$.
If X is of index one and Y is of index two, then, by the result quoted in the beginning of this section, C is a line or a conic.
If C is a smooth conic, we look at the generic isomorphism

$$\phi : (\mathcal{I}_C/\mathcal{I}_C^2)^* \to (\mathcal{I}_D/\mathcal{I}_D^2)^*|_C = \mathcal{O}_C(m) \oplus \mathcal{O}_C(-m).$$

Immediately we get that m is equal to one or two. Suppose $m = 2$. Then, by the Lemma, X is a quartic and there is a plane P tangent to X along C. Choose the coordinates so that P is given by $x_3 = x_4 = 0$. Then the equation of X can be written as

$$(q(x_0, x_1, x_2))^2 + x_3F + x_4G = 0,$$

where q defines C and F, G are cubic polynomials. Denote as A and B the curves cut out on P by these cubics. The necessary condition for smoothness of X is

$$A \cap B \cap X = \emptyset.$$

Now recall that C resp. P varies in a one-dimensional (complete) family C_t resp. P_t. A and B also vary, and for every t we must have

$$A_t \cap B_t \cap X = \emptyset.$$

This means that all the planes P_t pass through the same point, not lying on X. Projecting from this point, we see that the surface W formed by our conics C_t is in the ramification locus of this projection. The Hurwitz formula then gives $W \in |\mathcal{O}_X(i)|$ with $i \leq 3$. Now Y is, by assumption, a cubic or an intersection of two quadrics. But then, as we saw, the surface U_Y of (-1,1)-lines is at least $5H_Y$, and an elementary calculation shows that it is impossible that the inverse image of the surface of (-1,1)-lines U_Y consists only from W and the ramification.
If C is a line, then the argument is similar. One only needs to prove the following
Claim:In this situation, if $m = 2$, the scheme D has another reduced irreducible component C_1, which intersects C.
Then of course either C_1, or $C \bigcup C_1$ is a conic, and one can proceed as above. The proof of this claim is elementary algebra. We will sketch it after finishing the following last step of the Theorem:
If X and Y are both of index one, we have that the inverse image of a line l on Y should consist of lines and conics; for C as above, we have a map

$$\phi : (\mathcal{I}_C/\mathcal{I}_C^2)^* \to \mathcal{O}_C \oplus \mathcal{O}_C(-m),$$

if l is (0,-1), or

$$\phi' : (\mathcal{I}_C/\mathcal{I}_C^2)^* \to \mathcal{O}_C(m) \oplus \mathcal{O}_C(-2m),$$

if l is (1,-2). As these maps must be generic isomorphisms, we get that in both cases $m = 1$, whether C is a conic or a line.

Proof of the claim: Notice that C must be (1,-2)-line. The cokernel of the natural map

$$\beta : \mathcal{I}_D/\mathcal{I}_D^2|_C \to \mathcal{I}_C/\mathcal{I}_C^2$$

is the sheaf $\mathcal{I}_{C,D}/\mathcal{I}_{C,D}^2$, supported on intersection points of C and other components of D. We see from our assumptions that it must have length one (so be supported at one point x). Suppose that C intersects non-reduced components of D at x. Let A be a local ring of D at x and $\mathfrak{p} \subset A$ a fiber of $\mathcal{I}_{C,D}$. Of course $\mathfrak{p}/\mathfrak{p}^2 \neq 0$ by Nakayama. To see that $dim\mathfrak{p}/\mathfrak{p}^2 \geq 2$, we find an ideal $\mathfrak{a} \subset \mathfrak{p}$, not contained in $\mathfrak{p}^2$. For example, we can take an ideal defining the union of C and the reduction of an irreducible but non-reduced component of D intersecting C. We have a surjection

$$\mathfrak{p}/\mathfrak{p}^2 \to (\mathfrak{p}/\mathfrak{a})/(\mathfrak{p}^2/(\mathfrak{p}^2 \cap \mathfrak{a})) \to 0,$$

which has non-trivial (again by Nakayama) image and non-trivial kernel, q. e. d..

Corollary 3.3 *Let X, Y be Fano threefolds of index one as in Theorem 3.1 i). Then any map between X and Y is an isomorphism.*

Proof: Iskovskih computed the third Betti numbers of all Fano threefolds (see also [M]), and in fact as soon as $deg(X) > deg(Y)$, then $b_3(X) < b_3(Y)$, so a morphism $f : X \to Y$ cannot exist.

Remark C: Some part of the argument of Theorem 3.1 goes through without assumptions on the very ampleness of the generator H of the Picard group. For example, when X is a quartic double solid, which is a Fano threefold of index two, all the "lines" C on X except possibly a finite number, have either trivial normal bundle, or the normal bundle $\mathcal{O}_C(H) \oplus \mathcal{O}_C(-H)$ (in other words, the surface which parametrizes lines on X, has only isolated singularities). One can then replace the words "$T_X(H)$ is globally generated", which are not true in general, by some "normal bundle arguments" as in the above proof. The same should hold for the Veronese double cone. See [W], [T] for details. As for maps *to* the quartic double solid, the argument goes through without changes.

Examples: Any cubic in $\mathbf{P}^4$ satisfies the assumption we made on Y. By our Theorem 3.1 , we get that if a Fano threefold X of index one with cyclic Picard group is mapped onto a cubic, then the degree of this map can only be $\frac{degX}{3}$. So if X admits such a map, then $deg(X)$ is divisible by 3. Of course there are Fano threefolds of index one which admit a map onto a cubic: we can take an intersection of a cubic cone and a quadric in $\mathbf{P}^5$. Theorem 3.1 shows that if a smooth complete intersection of type (2,3) in $\mathbf{P}^5$ maps to a cubic, then it is contained in a cubic cone and the map is the projection from the vertex of this cone.
The same applies of course to maps from a complete intersection of three quadrics in $\mathbf{P}^6$ to a complete intersection of two quadrics in $\mathbf{P}^5$. Notice that any smooth complete intersection of two quadrics in $\mathbf{P}^5$ admits a map g onto a quadric in $\mathbf{P}^4$ such that the inverse image of the hyperplane section is the hyperplane section (any pencil of quadrics with non-singular base locus contains a quadratic cone). Therefore if a smooth intersection of three quadrics in $\mathbf{P}^6$ can be mapped onto a smooth complete

intersection of two quadrics in $\mathbf{P}^5$, it must lie in a quadric of corank 2 in $\mathbf{P}^6$. Of course a general intersection of three quadrics in $\mathbf{P}^6$ does not have this property, as the space of quadrics of corank 2 is of codimension four in the space of all quadrics.

ADDITIONAL EXAMPLES OF VARIETIES SATISFYING THE ASSUMPTION OF THEOREM 3.1:

1) any complete intersection of a cubic and a quadric in $\mathbf{P}^5$ or
2) any complete intersection of three quadrics in $\mathbf{P}^6$. Indeed, if lines on these varieties cover only a hyperplane section divisor, then the scheme of lines must be non-reduced, i.e. each line must have normal bundle $\mathcal{O}_{\mathbf{P}^1}(-2) \oplus \mathcal{O}_{\mathbf{P}^1}(1)$. So the surface of lines is either a cone or the tangent surface to a curve. But one can check that these varieties do not contain cones; neither do they contain a tangent surface to a curve as a hyperplane section, because by a version of Zak's theorem on tangencies (see for example [FL]), a hyperplane section of a complete intersection has only isolated singularities.
3) Any V_{22} with reduced Hilbert scheme of lines. By ([P]), there is exactly one V_{22} such that its Hilbert scheme of lines is non-reduced.
4) any Fano threefold V_{16} of index one and genus 9. This can be shown by the method of Prokhorov ([P]) :
First, notice that if the lines on V_{16} cover only a hyperplane section, the scheme of lines is non-reduced. So all the lines are tangent to a curve. This is actually a rational normal curve, so the lines never intersect.
For convenience of the reader, we recall from [I2] the notion of double projection from a line and its application to V_{16} :
Let X be a Fano threefold of index one, $g(X) \geq 7$, and let l be a line on X. On $\tilde{X}$, the blow-up of X, we consider the linear system $|\sigma^*H - 2E|$, where σ is the blow-up, $H = K_Y$ and E is the exceptional divisor. This is not base-point-free, namely, its base locus consists of proper preimages of lines intersecting l, and, if l is (-2,1), from the exceptional section of the ruled surface $E \cong F_3$. However, after a flop (i.e. a birational transformation which is an isomorphism outside this locus) we can make it into a base-point-free system $|(\sigma^*H)^+ - 2E^+|$ on the variety $\tilde{X}^+$.
If $g(X) = 9$, i.e.X is a V_{16}, the variety $\tilde{X}^+$ is birationally mapped by this linear system onto $\mathbf{P}^3$. This map, say g, is a blow-down of the surface of conics intersecting l to a curve $Y \subset \mathbf{P}^3$, which is smooth of degree 7 and genus three (smoothness of Y is obtained from Mori's extremal contraction theory). Y lies on a cubic surface which is the image of E^+. Moreover, the inverse rational map from $\mathbf{P}^3$ to X is given by the linear system $|7H - 2Y|$.
One has therefore that the lines from X, different from l, must be mapped by g to trisecants of Y. Note that if lines on X form only a hyperplane section, the desingularization of the surface of lines on X is rational ruled, and it remains so after the blow-up and the flop. So, as in [P], we must have a morphism $F_e \to \mathbf{P}^3$, which is given by some linear system $|C + kF|$ with C the canonical section and f a fiber, such that the inverse image of Y belongs to the system $|3C + lF|$. $deg(Y) = 7$ implies

$$(3C + lF)(C + kF) = -3e + 3k + l = 7,$$

and as $degK_Y = 4$,

$$(C + (l - 2 - e)F)(3C + lF) = -6e + 4l - 6 = 4,$$

Combining these two equations, we get

$$2k - e = 3,$$

However, we must have $e \geq 0$ and $k \geq e$, as otherwise the linear system $|C + kF|$ does not define a morphism. This leaves only two possibilities for k and e: either $e = k = 3$, or $e = 1, k = 2$. The first case actually cannot occur: this would imply that Y is singular. So the image of $F_e = F_1$ in $\mathbf{P}^3$ is a cubic which is a projection of F_1 from $\mathbf{P}^4$. By assumption, Y is also contained in another irreducible cubic (the image of E^+). But one check that this cannot happen, using e.g. a theorem by d'Almeida ([Al]), which asserts that if a smooth non-degenerate curve Y of degree $d \geq 6$ and genus g in $\mathbf{P}^3$ satisfies $H^1(\mathcal{I}_Y(d-4)) \neq 0$, then Y has a (d-2)-secant provided that $(d, g) \neq (7, 0), (7, 1), (8, 0)$.

4. V_{22}

Let us now take $Y = V_{22}^s$, i.e. the only variety of type V_{22} which has non-reduced Hilbert scheme of lines. This V_{22} violates the assumptions of Theorem 3.1. However, using Mukai's and Schreyer's descriptions of conics on varieties of type V_{22}, it is still possible to say something on maps from Fano threefolds onto Y. We will show the following:

Proposition 4.1 *A Fano threefold X of index two with cyclic Picard group and irreducible Hilbert scheme of lines does not admit a map onto V_{22}^s.*

As for the last assumption on X, one believes that this is always satisfied. In fact this is easy to check (and well-known) for a cubic or a complete intersection of two quadrics (the Hilbert scheme is smooth in this case, so it is enough to show that it is connected). The irreducibility is also known for V_5, in fact, the Hilbert scheme is isomorphic to $\mathbf{P}^2$ ([I], I, Corollary 6.6). For a quartic double solid, see [W]. As for a double Veronese cone, in [T] it is proven that a general double Veronese cone has irreducible Hilbert scheme of lines. So the only possible exception could be a special double Veronese cone.

In fact our argument will work for a sufficiently general V_{22}, but for all of them except V_{22}^s this assertion is already proved in the last paragraph.

Proof: Let S be the Fano surface (= reduced Hilbert scheme) of lines on X and T the Fano surface of conics on the V_{22}. If $f : X \to V_{22}$ is a finite map, then, as Schuhmann proves in [S], the inverse image of any conic is a union of lines, and, moreover, in this way f induces a finite surjective morphism $g : S \to T$ (thanks to irreducibility of S, any line on X is in the inverse image of a conic on V_{22}).
F.-O. Schreyer ([Sch]) gives the following description of a general conic on V_{22}:
Consider V_{22} as the variety of polar hexagons of a plane quartic curve $C \subset \mathbf{P}^2$ (a polar hexagon of C is the union of six lines $l_1, ...l_6$ given by equations $L_1 = 0, ..., L_6 = 0$,

such that $L_1^4 + ... + L_6^4 = F$ where $F = 0$ defines C; "the variety of polar hexagons" means here the closure of the set of 6-tuples $l_1, ...l_6$ with $L_1^4 + ... + L_6^4 = F$ in the Hilbert scheme $Hilb_6(\mathbf{P}^{2*})$; a general V_{22} is isomorphic to such a variety for a certain curve C; V_{22}^s is the variety of polar hexagons of a double conic). Then there is a birational isomorphism between $(\mathbf{P}^2)^*$ and T given as follows:
for a general $l \subset \mathbf{P}^2$ the curve of polar hexagons to C containing l is a conic on V_{22}.
This description and the fact that through any point on a V_{22} there is only a finite number of conics ([I], II, Theorem 4.4) gives that
there are six conics through a general point of V_{22}.
In [M], Mukai claims that the Fano surface of conics on a V_{22} is even isomorphic to $\mathbf{P}^2$. Unfortunately, this paper does not contain a proof of this fact. The proof appears in the paper of A. Kuznetsov ([K]): he uses another description of a general V_{22} as a subvariety of $G(2,6)$. Namely, if V and N are 7- and 3-dimensional vector spaces respectfully and $f : N \to \Lambda^2 V^*$ is a general net of skew-symmetric forms on V, then a general V_{22} (including V_{22}^s, [Sch]) appears as a set of all 3-subspaces of V which are isotropic with respect to this net (i.e. to all forms of the net simultaneously). Let U (resp. Q) denote restriction on a V_{22} of the universal (resp. universal quotient) bundle on $G(2,6)$. Kuznetsov proves that every (possibly singular) conic on a V_{22} is a degeneracy locus of a homomorphism $U \to Q^*$; the Fano surface of conics is thus $\mathbf{P}(Hom(U, Q^*)) = \mathbf{P}^2$.
Now if there is a finite map $f : X \to V_{22}$ as above, then X must be a cubic: indeed, a Fano threefold with cyclic Picard group and with 6 lines through a general point is a cubic. Let $f^* H_{V_{22}} = mH_X$, then one easily computes that the inverse image of a general conic consists of $deg(g) = s = \frac{3}{11}m^2$ lines.
For simplicity, we will use the same notation for points of T (resp. S) and corresponding conics on V_{22} (resp. lines on X). We have $T \cong \mathbf{P}^2$. Let a be such that conics on V_{22} intersecting a given (general) conic A, form a divisor D_A from $|\mathcal{O}_{\mathbf{P}^2}(a)|$
On S, denote as E_L the divisor of lines intersecting a given line L. It is well-known and easy to compute that $E_L \cdot E_M = 5$ for any L, M.
If $g^{-1}(A) = \{L_1, ..., L_s\}$, then

$$g^*(\mathcal{O}_{\mathbf{P}^2}(a)) = \mathcal{O}_S(E_{L_1} + ... + E_{L_s}).$$

We therefore have another formula for $deg(g)$:

$$deg(g) = \frac{5s^2}{a^2}.$$

From the equality $s = \frac{5s^2}{a^2}$ we get that $(\frac{m}{a})^2 = \frac{11}{15}$, however, this is impossible as $\frac{11}{15}$ is not a square of a rational number.

References

[Al] J. d'Almeida: Courbes de l'espace projectif: Series lineaires incompletes et multisecantes, J. Reine Angew. Math., 370 (1986), 30-51.

[A] E. Amerik: On a problem of Noether-Lefschetz type, to appear in Compositio Mathematica.

[AW] M. Andreatta, J. Wisniewski: On contractions of smooth varieties, MSRI e-print archive, preprint alg-geom/9605013

[B] R. Braun: On the normal bundles of Cartier divisors on projective varieties, Arch. Math. 59 (1992), 403-411.

[CG] H. Clemens, Ph. Griffiths: The intermediate Jacobian of the cubic threefold, Ann. Math. 95 (1972), 281-356.

[D] J.-P. Demailly: L^2-vanishing theorems for positive line bundles and adjunction theory, Cetraro Lectures, 1994, MSRI e-prints archive, preprint alg-geom/9410022.

[EGPS] G. Ellingsrud, L. Gruson, C. Peskine and S. Stromme: On the normal bundle of curves on smooth projective surfaces, Invent. Math. 80 (1985), 181-184.

[FL] W.Fulton, R. Lazarsfeld: Connectivity and its applications in algebraic geometry, in *Algebraic geometry, Proceedings, University of Illinois at Chicago Circle, 1980*, Lecture Notes in Math. 862.

[G] Ph. Griffiths: Hermitian differential geometry, Chern classes and vector bundles, in Global Analysis, papers in honor of K. Kodaira, Princeton University press, 1969, 185-251.

[GH] Ph. Griffiths, J. Harris: *Principles of algebraic geometry*, Wiley, 1978.

[I] V.A. Iskovskih: Fano threefolds I, II, Math. USSR Izv. 11 (1977), 485-527, and 12 (1978), 469-506.

[I1] V. A. Iskovskih: Anticanonical models of 3-dimensional algebraic varieties, in: R.V. Gamkrelidze (ed.), Itogi nauki i techniki, Sovremennie problemi matematiki, t. 12 (in Russian).

[I2] V.A. Iskovskih: Double projection from a line on Fano threefolds of the first kind, Math. USSR Sbornik, 66 (1990), no. 1, 265-284.

[K] A. Kuznetsov: "The derived category of coherent sheaves on V_{22}, master's thesis, Moscow, 1995.

[M] S. Mukai: Fano 3-folds, in Complex Projective Geometry, ed. by G. Ellingsrud et al., London Math. Soc. Lecture Notes 179, 255-263.

[P] Yu. Prokhorov: On exotic Fano varieties, Moscow Univ. Math. Bull. 45 (1990), no.3, 36-38.

[Sch] F.-O. Schreyer: Geometry and algebra of prime Fano threefolds of genus 12, preprint (preliminary partial version).

[S] C. Schuhmann: Morphisms between Fano threefolds, thesis, Leiden, 1997.

[T] A. Tikhomirov: The Fano surface of lines on the Veronese double cone, Math. USSR Izv. 19 (1982), 377-443.

[W] G. Welters: Abel-Jacobi isogenies for certain types of Fano threefolds, thesis, Utrecht, 1981.

Ekaterina Amerik
Institut Fourier
100 Rue des Maths
BP 74 38402 St.-Martin d'Heres
France
amerik@main.mccme.rssi.ru

Invariant Inner Product in Spaces of Holomorphic Functions on Bounded Symmetric Domains

Jonathan Arazy and Harald Upmeier[1]

Received: March, 24, 1996
Revised: August, 8, 1997

Communicated by Joachim Cuntz

Abstract. We provide new integral formulas for the invariant inner product on spaces of holomorphic functions on bounded symmetric domains of tube type.

1991 Mathematics Subject Classification: Primary 46E20; Secondary 32H10, 32A37, 43A85

0 Introduction

Our main concern in this work is to provide *concrete formulas for the invariant inner products and hermitian forms on spaces of holomorphic functions on Cartan domains* D *of tube type*. As will be explained below, the group $Aut(D)$ of all holomorphic automorphisms of D acts transitively. $Aut(D)$ acts projectively on function spaces on D via $f \mapsto U^{(\lambda)}(\varphi)f := (f \circ \varphi)\,(J\varphi)^{\lambda/p}, \quad \varphi \in Aut(D), \ \lambda \in \mathbf{C}$, but these actions are not irreducible in general. The inner products we consider are those obtained from the holomorphic discrete series by analytic continuation. The associated Hilbert spaces generalize the weighted Bergman spaces, the Hardy and the Dirichlet space. By "concrete" formulas we mean Besov-type formulas, namely integral formulas involving the functions and some of their derivatives. Possible applications include the study of operators (Toeplitz, Hankel) acting on function spaces and the theory of invariant Banach spaces of analytic functions (where the pairing between an invariant space and its invariant dual is computed via the corresponding invariant inner product).

Our problem is closely related to finding concrete realizations (by means of integral formulas) of the analytic continuation of the *Riesz distribution.* [Ri], [Go], [FK2], Chapter VII.

[1]Authors supported
by a grant from the German-Israeli Foundation (GIF), I-415-023.06/95.

In principle, the analytic continuation is obtained from the integral formulas associated with the weighted Bergman spaces (i.e. the holomorphic discrete series) by "partial integration with respect to the radial variables". This program has been successful in the case of rank 1 (i.e. when D is the open unit ball of $\mathbf{C}^d$, see [A3]). The case of rank $r > 1$ is more difficult, and concrete formulas are known only in special cases, see [A2], [Y4], [Y1], [Y2].

This paper consists of two main parts. In the first part (Sections 2, 3, and 4) we develop in full generality the techniques of [A2], [Y4], and obtain integral formulas for the invariant inner products associated with the so-called Wallach set and pole set. In the second part (section 5) we introduce new techniques (integration on boundary orbits), to obtain new integral formulas for the invariant inner products in the important special cases of Cartan domains of type I and IV. This approach has the potential for further generalizations and applications, including the infinite dimensional setup.

The paper is organized as follows. Section 1 provides background information on Cartan domains, the associated symmetric cones and Siegel domains and the Jordan theoretic approach to the study of bounded symmetric domains [Lo], [FK2], [U2]. We also explain some general facts concerning invariant Hilbert spaces of analytic functions on Cartan domains and the connection to the Riesz distribution. Section 2 is devoted to the study of invariant differential operators on symmetric cones. We study the "shifting operators" introduced by Z. Yan and their derivatives with respect to the "spectral parameter". Section 3 is devoted to our generalization of Yan's shifting method, to find explicit integral formulas for the invariant inner products obtained by analytic continuation of the holomorphic discrete series. In section 4 we study the expansion of Yan's operators, and obtain applications to concrete integral formulas for the invariant inner products. Some of these results were obtained independently by Z. Yan, J. Faraut and A. Korányi, [FK2], [Y4]. We include these results and our proofs, in order to make the paper self contained, and also because in most cases our results go beyond the results in [FK2], [Y4].

In section 5 we propose a new type of integral formulas for the invariant inner products. These formulas involve integration on boundary orbits and the application of the localized versions of the radial derivative associated with the boundary components of Cartan domains. We are able to establish the desired formulas in the important special cases of type I and IV. The techniques established in this section can be used in the study of the remaining cases.

Finally, in the short section 6 we use the quasi-invariant measures on the boundary orbits of the associated symmetric cone in order to obtain integral formulas for some of the invariant inner products in the context of the *unbounded* realization of the Cartan domains (tube domains). These results are essentially implicitly contained in [VR], where the authors use the Lie-theoretic and Fourier-analytic approach to analysis on symmetric Siegel domains. We use the Jordan-theoretic approach which yields simpler formulation of the results and simpler proofs.

ACKNOWLEDGMENT: We would like to thank Z. Yan, J. Faraut, and A. Korányi for sending us drafts of their work and for many stimulating discussions. We also thank the referee for valuable comments on the first version of the paper.

1 Preliminaries

A *Cartan domain* $D \subset \mathbf{C}^d$ is an irreducible bounded symmetric domain in its Harish-Chandra realization. Thus D is the open unit ball of a Banach space $Z = (\mathbf{C}^d, \|\cdot\|)$ which admits the structure of a *JB*-triple*, namely there exists a continuous mapping $Z \times Z \times Z \ni (x,y,z) \to \{x,y,z\} \in Z$ (called the *Jordan triple product*) which is bilinear and symmetric in x and z, conjugate-linear in y, and so that the operators $D(x,x): Z \to Z$ defined for every $x \in Z$ by $D(x,x)z := \{x,x,z\}$ are hermitian, have positive spectrum, satisfy the "C*-axiom" $\|D(x,x)\| = \|x\|^2$, and the operators $\delta(x) := iD(x,x)$ are *triple derivations*, i.e. the *Jordan triple identity holds*

$$\delta(x)\{y,z,w\} = \{\delta(x)y,z,w\} + \{y,\delta(x)z,w\} + \{y,z,\delta(x)w\}, \quad \forall y,z,w \in Z.$$

The norm $\|\cdot\|$ is called the *spectral norm*. We put also $D(x,y)z := \{x,y,z\}$. An element $v \in Z$ is called a *tripotent* if $\{v,v,v\} = v$. Every tripotent $v \in Z$ gives rise to a direct-sum *Peirce decomposition*

$$Z = Z_1(v) + Z_{\frac{1}{2}}(v) + Z_0(v), \quad \text{where} \quad Z_\nu(v) := \{z \in Z;\ D(v,v)z = \nu z\}, \quad \nu = 1, \frac{1}{2}, 0.$$

The associated *Peirce projections* are defined for $z_\kappa \in Z_\kappa(v)$, $\kappa = 1, \frac{1}{2}, 0$, by

$$P_\nu(v)(z_1 + z_{\frac{1}{2}} + z_0) = z_\nu, \quad \nu = 1, \frac{1}{2}, 0.$$

In this paper we are interested in the important special case where Z contains a *unitary* tripotent e, for which $Z = Z_1(e)$. In this case Z has the structure of a *JB*-algebra* with respect to the binary product $x \circ y := \{x,e,y\}$ and the involution $z^* := \{e,z,e\}$, and e is the unit of Z. The binary Jordan product is commutative, but in general non-associative. The triple product is expressed in terms of the binary product and the involution via $\{x,y,z\} = (x \circ y^*) \circ z + (z \circ y^*) \circ x - (x \circ z) \circ y^*$. In this case the open unit ball D of Z is a *Cartan domain of tube-type*. This terminology is related to the *unbounded realization of D*, to be explained later.

Let $X := \{x \in Z; x^* = x\}$ be the real part of Z. It is a formally-real (or euclidean) Jordan algebra. Every $x \in X$ has a *spectral decomposition* $x = \sum_{j=1}^r \lambda_j e_j$, where $\{e_j\}_{j=1}^r$ is a *frame* of pairwise orthogonal minimal idempotents in X, and $\{\lambda_j\}_{j=1}^r$ are real numbers called the *eigenvalues* of x. The *trace* and *determinant* (or, "norm") are defined in X via

$$tr(x) := \sum_{j=1}^r \lambda_j, \qquad N(x) := \prod_{j=1}^r \lambda_j$$

respectively, and they are polynomials on X. The maximal number r of pairwise orthogonal minimal idempotents in X is called the *rank* of X. The positive-definite inner product in X, $\langle x,y\rangle = tr(x \circ y)$, $\quad x,y \in X$, satisfies

$$\langle x \circ y, z\rangle = \langle x, y \circ z\rangle, \quad x,y,z \in X.$$

Equivalently, the multiplication operators $L(x)y := x \circ y$, $\quad x,y \in X$, are self-adjoint. The trace and determinant polynomials, as well as the multiplication operators, have unique extensions to the complexification $X^{\mathbf{C}} := X + iX = Z$. Let

$$\Omega := \{x^2; x \in X, N(x) \neq 0\}.$$

Then Ω is a *symmetric, open convex cone*, i.e. Ω is *self polar* and *homogeneous* with respect to the group $GL(\Omega)$ of linear automorphisms of Ω. We denote the connected component of the identity in $GL(\Omega)$ by $G(\Omega)$. Define

$$P(x) := 2L(x)^2 - L(x^2), \quad x \in X, \tag{1.1}$$

then $P(x) \in G(\Omega)$ for every $x \in \Omega$, and $x = P(x^{1/2})e$. Thus $G(\Omega)$ is transitive on Ω. The map $x \to P(x)$ from X into $End(X)$ is called the *quadratic representation* because of the identity

$$P(P(x)y) = P(x)P(y)P(x), \quad \forall x, y \in X. \tag{1.2}$$

The domain $T(\Omega) := X + i\Omega$, called the *tube over* Ω. It is an irreducible symmetric domain which is biholomorphically equivalent to D by means of the *Cayley transform* $c : D \to T(\Omega)$, defined by

$$c(z) := i\frac{e+z}{e-z}, \quad z \in Z.$$

This explains why D is called a tube-type Cartan domain.

Let $e_1, e_2, \dots, e_r$ be a fixed frame of minimal, pairwise orthogonal idempotents satisfying $e_1 + e_2 + \dots + e_r = e$, where e is the unit of Z. Let

$$Z = \sum_{1 \le i \le j \le r} Z_{i,j}$$

be the associated *joint Peirce decomposition*, namely $Z_{i,j} := Z_{\frac{1}{2}}(e_i) \cap Z_{\frac{1}{2}}(e_j)$ for $1 \le i < j \le r$ and $Z_{i,i} := Z_1(e_i)$ for $1 \le i \le r$. The *characteristic multiplicity* is the common dimension $a = \dim(Z_{i,j})$, $1 \le i < j \le r$, and $d = r + r(r-1)a/2$. The number $p := (r-1)a + 2$ is called the *genus* of D. It is known that

$$Det(P(x)) = N(x)^p, \quad \forall x \in X,$$

where "Det" is the usual determinant polynomial in $End(Z)$. From this and (1.2) it follows that

$$N(P(x)y) = N(x)^2 N(y) \quad \forall x, y \in X. \tag{1.3}$$

Let $u_j := e_1 + e_2 + \dots + e_j$ and let $Z_j := \sum_{1 \le i \le k \le j} Z_{i,k}$ be the JB*- subalgebra of Z whose unit is u_j. Let N_j be the determinant polynomials of the $Z_j, 1 \le j \le r$; they are called the *principal minors* associated with the frame $\{e_j\}_{j=1}^r$. Notice that $Z_r = Z$ and $N_r = N$. For an r-tuple of integers $\mathbf{m} = (m_1, m_2, \dots, m_r)$ write $\mathbf{m} \ge 0$ if $m_1 \ge m_2 \ge \dots \ge m_r \ge 0$. Such r-tuples $\mathbf{m}$ are called *signatures* (or, "partitions"). The *conical polynomial* associated with the signature $\mathbf{m}$ is

$$N_{\mathbf{m}}(z) := N_1(z)^{m_1 - m_2} N_2(z)^{m_2 - m_3} N_3(z)^{m_3 - m_4} \dots N_r(z)^{m_r}, \quad z \in Z.$$

Notice that $N_{\mathbf{m}}(\sum_{j=1}^r t_j e_j) = \prod_{j=0}^r t_j^{m_j}$, thus the conical polynomials are natural generalizations of the monomials. Let $Aut(D)$ be the group of all biholomorphic automorphisms of D, and let G be its connected component of the identity. Let $K := \{g \in G; g(0) = 0\} = G \cap GL(Z)$ be the maximal compact subgroup of G. For any signature $\mathbf{m}$ let $P_{\mathbf{m}} := span\{N_{\mathbf{m}} \circ k; k \in K\}$. Clearly, $P_{\mathbf{m}} \subset \mathcal{P}_\ell$, where

$\ell = |\mathbf{m}| = \sum_{j=1}^{r} m_j$ and $\mathcal{P}_\ell$ is the space of homogeneous polynomials of degree ℓ. By definition, $P_{\mathbf{m}}$ are invariant under the composition with members of K. Let

$$\langle f, g\rangle_{\mathcal{F}} := \partial_f(g^\sharp)(0) = \frac{1}{\pi^d}\int_Z f(z)\overline{g(z)}\, e^{-|z|^2} dm(z) \tag{1.4}$$

be the *Fock-Fischer inner product* on the space $\mathcal{P}$ of polynomials, where $g^\sharp(z) := \overline{g(z^*)}$, $\partial_f = f(\frac{\partial}{\partial z})$, $|z|$ is the unique K-invariant Euclidean norm on Z normalized so that $|e_1| = 1$, and $dm(z)$ is the corresponding Lebesgue volume measure. (Thus $\langle 1, 1\rangle_{\mathcal{F}} = 1$). The following result (*Peter-Weyl decomposition*) is proved in [Sc], see also [U1]. Here the group K acts on functions on D via $\pi(k)f := f \circ k^{-1}$, $k \in K$. Notice that $\mathcal{P}_\ell$, $\ell = 0, 1, 2, \ldots$ and $\mathcal{P}$ are invariant under this action.

THEOREM 1.1 (I) *The spaces $\{P_{\mathbf{m}}\}_{\mathbf{m}\geq 0}$, are K-invariant and irreducible. The representations of K on the spaces $P_{\mathbf{m}}$ are mutually inequivalent, the $P_{\mathbf{m}}$'s are mutually orthogonal with respect to $\langle\cdot,\cdot\rangle_{\mathcal{F}}$, and $\mathcal{P} = \sum_{\mathbf{m}\geq 0} P_{\mathbf{m}}$.*
(II) *If $\mathcal{H}$ is a Hilbert space of analytic functions on D with a K-invariant inner product in which the polynomials are dense, then $\mathcal{H}$ is the orthogonal direct sum $\mathcal{H} = \sum_{\mathbf{m}\geq 0} \oplus P_{\mathbf{m}}$. Namely, every $f \in \mathcal{H}$ is expanded in the norm convergent series $f = \sum_{\mathbf{m}\geq 0} f_{\mathbf{m}}$, with $f_{\mathbf{m}} \in P_{\mathbf{m}}$, and the spaces $P_{\mathbf{m}}$ are mutually orthogonal in $\mathcal{H}$. Moreover, there exist positive numbers $\{c_{\mathbf{m}}\}_{\mathbf{m}\geq 0}$ so that for every $f, g \in \mathcal{H}$ with expansions $f = \sum_{\mathbf{m}\geq 0} f_{\mathbf{m}}$ and $g = \sum_{\mathbf{m}\geq 0} g_{\mathbf{m}}$ we have*

$$\langle f, g\rangle_{\mathcal{H}} = \sum_{\mathbf{m}\geq 0} c_{\mathbf{m}} \langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}}.$$

For every signature $\mathbf{m}$ let $K_{\mathbf{m}}(z, w)$ be the reproducing kernel of $P_{\mathbf{m}}$ with respect to (1.4). Clearly, the reproducing kernel of the *Fock-Fischer space* $\mathcal{F}$ (the completion of $\mathcal{P}$ with respect to $\langle\cdot,\cdot\rangle_{\mathcal{F}}$) is

$$F(z, w) := \sum_{\mathbf{m}} K_{\mathbf{m}}(z, w) = e^{\langle z, w\rangle}.$$

An important property of the norm polynomial N is its transformation rule under the group K

$$N(k(z)) = \chi(k)N(z), \quad k \in K, \ z \in Z \tag{1.5}$$

where $\chi : K \to \mathbf{T} := \{\lambda \in \mathbf{C}; |\lambda| = 1\}$ is a character. In fact, $\chi(k) = N(k(e)) = Det(k)^{2/p}$ $\forall k \in K$. Notice that (1.5) implies that $P_{(m,m,\ldots,m)} = \mathbf{C}\, N^m$ for $m = 0, 1, 2, \ldots$.

The subgroup L of K defined via

$$L := \{k \in K; k(e) = 1\} \tag{1.6}$$

plays an important role in the theory.

LEMMA 1.1 *For every signature $\mathbf{m} \geq 0$ the function*

$$\phi_{\mathbf{m}}(z) := \int_L N_{\mathbf{m}}(\ell(z))d\ell \tag{1.7}$$

is the unique spherical (i.e., L-invariant) polynomial in $P_{\mathbf{m}}$ satisfying $\phi_{\mathbf{m}}(e) = 1$.

For example, $\phi_{(m,m,\ldots,m)} = N^m$ by (1.5). The L-invariant real polynomial on X

$$h(x) = h(x,x) := N(e - x^2)$$

admits a unique K-invariant, hermitian extension $h(z,w)$ to all of Z. Thus, $h(k(z), k(w)) = h(z,w)$ for all $z, w \in Z$ and $k \in K$, $h(z,w)$ is holomorphic in z and anti-holomorphic in w, and $h(z,w) = \overline{h(w,z)}$, [FK1]. The transformation rule of $h(z,w)$ under $Aut(D)$ is

$$h(\varphi(z), \varphi(w)) = J\varphi(z)^{\frac{1}{p}}\ h(z,w)\ \overline{J\varphi(w)}^{\frac{1}{p}}, \quad \varphi \in Aut(D),\ z, w \in D, \tag{1.8}$$

where $J\varphi(z) := Det(\varphi'(z))$ is the complex Jacobian of φ, and p is the genus of D.

For $\mathbf{s} = (s_1, s_2, \ldots, s_r) \in \mathbf{C}^r$ one defines the *conical function* $N_{\mathbf{s}}$ on Ω via

$$N_{\mathbf{s}}(x) := N_1^{s_1 - s_2}(x)\, N_2^{s_2 - s_3}(x)\, N_3^{s_3 - s_4}(x) \ldots \cdot N_r^{s_r}(x), \quad x \in \Omega,$$

which generalize the conical polynomials $N_{\mathbf{m}}$. In what follows use the following notation:

$$\lambda_j := (j-1)\frac{a}{2}, \quad 1 \le j \le r.$$

The *Gindikin - Koecher Gamma function* is defined for $\mathbf{s} = (s_1, s_2, \ldots, s_r) \in \mathbf{C}^r$ with $\Re(s_j) > \lambda_j$, $1 \le j \le r$, via

$$\Gamma_\Omega(\mathbf{s}) := \int_\Omega e^{-tr(x)} N_{\mathbf{s}}(x) d\mu_\Omega(x).$$

Here $tr(x) = \langle x, e\rangle$ is the Jordan-theoretic trace of x, and

$$d\mu_\Omega(x) := N(x)^{-\frac{d}{r}} dx$$

is the (unique, up to a multiplicative constant) $G(\Omega)$-*invariant measure* on Ω. The following formula [Gi] reduces the computation of $\Gamma_\Omega(\mathbf{s})$ to that of ordinary Gamma functions:

$$\Gamma_\Omega(\mathbf{s}) = (2\pi)^{(d-r)/2} \prod_{1 \le j \le r} \Gamma(s_j - \lambda_j), \tag{1.9}$$

and provides a meromorphic continuation of Γ_Ω to all of $\mathbf{C}^r$. In particular, $\Gamma_\Omega(\lambda) := \Gamma_\Omega(\lambda, \lambda, \ldots, \lambda)$ is given by

$$\Gamma_\Omega(\lambda) = \int_\Omega e^{-tr(x)}\ N(x)^\lambda\ d\mu_\Omega(x) = (2\pi)^{(d-r)/2} \prod_{1 \le j \le r} \Gamma(\lambda - \lambda_j),$$

and it is an entire meromorphic function. The *pole set* of $\Gamma_\Omega(\lambda)$ is precisely

$$\mathbf{P}(D) := \cup_{1 \le j \le r}(\lambda_j - \mathbf{N}) = \{\lambda_j - n;\ 1 \le j \le r,\ n \in \mathbf{N}\}. \tag{1.10}$$

For $\lambda \in \mathbf{C}$ and a signature $\mathbf{m} = (m_1, m_2, \ldots, m_r)$ one defines

$$(\lambda)_{\mathbf{m}} := \frac{\Gamma_\Omega(\mathbf{m} + \lambda)}{\Gamma_\Omega(\lambda)} = \prod_{j=1}^r (\lambda - \lambda_j)_{m_j} = \prod_{j=1}^r \prod_{n=0}^{m_j - 1} (n + \lambda - \lambda_j),$$

where $\mathbf{m}+\lambda := (m_1+\lambda, m_2+\lambda, \dots, m_r+\lambda)$.

We recall two important formulas for *integration in polar coordinates* [FK2], Chapters VI and IX. The first formula uses the fact that $K\cdot\Omega = Z$, namely the fact that every $z\in Z$ can be written (not uniquely) in the form $z = k(x)$, where $x\in\Omega$ and $k\in K$. This is the first (or "conical") type of *polar decomposition* of x, and it generalizes the polar decomposition of matrices. This leads to the formula

$$\int_Z f(z)dm(z) = \frac{\pi^d}{\Gamma_\Omega(\frac{d}{r})}\int_\Omega\left(\int_K f(k(x^{\frac{1}{2}}))\,dk\right)\,dx \tag{1.11}$$

which holds for every $f\in L^1(Z,m)$. Next, fix a frame $e_1,\dots,e_r$, and define

$$R := \mathrm{span}_{\mathbf{R}}\{e_j\}_{j=1}^r \quad\text{and}\quad R_+ := \{\sum_{j=1}^r t_je_j;\ t_1>t_2>\ldots>t_r>0\}$$

and

$$\mathbf{R}_+^r := \{t=(t_1,\ldots t_r);\ t_1>t_2>\ldots>t_r>0\}.$$

Then $Z = K\cdot R$, namely every $z\in Z$ has a representation $z = k(\sum_{j=1}^r t_je_j)$ for some (again, not unique) $\sum_{j=1}^r t_je_j \in R$ and $k\in K$. This representation is the second type of polar decomposition of z. Moreover, $m(Z\setminus K\cdot R_+) = 0$, namely up to a subset of measure zero, every $z\in Z$ has a representation $z = k(\sum_{j=1}^r t_j^{1/2}e_j)$ with $t_1>t_2>\ldots>t_r>0$. This leads to the formula

$$\int_Z f(z)dm(z) = c_0\int_{\mathbf{R}_+^r}\left(\int_K f(k(\sum_{j=1}^r t_j^{\frac{1}{2}}e_j))\,dk\right)\prod_{1\le i<j\le r}(t_i-t_j)^a\,dt_1\,dt_2\ldots dt_r, \tag{1.12}$$

which holds for every $f\in L^1(Z,m)$. The constant c_0 will be determined as a by-product of our work in section 5 below. For convenience, we can write (1.12) in the form

$$\int_Z f(z)dm(z) = c_0\int_{\mathbf{R}_+^r} f^{\#}(t)\,w(t)^a\,dt, \tag{1.13}$$

where

$$f^{\#}(t) := \int_K f(k(\sum_{j=1}^r t_j^{\frac{1}{2}}e_j))\,dk,\quad t=(t_1,t_2,\ldots,t_r)\in\mathbf{R}_+^r$$

is the *radial part* of F and

$$w(t) := \prod_{1\le i<j\le r}(t_i-t_j),\quad t=(t_1,t_2,\ldots,t_r)\in\mathbf{R}_+^r \tag{1.14}$$

is the *Vandermonde polynomial.*

By [Hu], [Be], [La1], [FK1], we have the *binomial formula*:

THEOREM 1.2 *For $\lambda\in\mathbf{C}$ we have*

$$N(e-x)^{-\lambda} = \sum_{\mathbf{m}\ge 0}(\lambda)_{\mathbf{m}}\frac{\phi_{\mathbf{m}}(x)}{\|\phi_{\mathbf{m}}\|_{\mathcal{F}}^2},\quad \forall x\in\Omega\cap(e-\Omega), \tag{1.15}$$

and

$$h(z,w)^{-\lambda} = \sum_{\mathbf{m}\geq 0} (\lambda)_{\mathbf{m}} K_{\mathbf{m}}(z,w), \quad \forall z,w \in D. \tag{1.16}$$

The two series converge absolutely, (1.15) converges uniformly on compact subsets of $(\lambda, x) \in \mathbf{C} \times (\Omega \cap (e - \Omega))$*, and (1.16) converges uniformly on compact subsets of* $(\lambda, z, w) \in \mathbf{C} \times D \times D$.

In particular, it follows that for fixed $z, w \in D$, the function $\lambda \to h(z,w)^{-\lambda}$ is analytic in all of $\mathbf{C}$ (this can be proved also by showing that $h(z,w) \neq 0$ for $z, w \in D$).

The *Wallach set* of D, denoted by $\mathbf{W}(D)$, is the set of all $\lambda \in \mathbf{C}$ for which the function $(z,w) \to h(z,w)^{-\lambda}$ is non-negative definite in $D \times D$, namely

$$\sum_{i,j} a_i \overline{a}_j \, h(z_i, z_j)^{-\lambda} \geq 0$$

for all finite sequences $\{a_j\} \subseteq \mathbf{C}$ and $\{z_j\} \subseteq D$. For $\lambda \in \mathbf{W}(D)$ let $\mathcal{H}_\lambda$ be the completion of the linear span of the functions $\{h(\cdot, w)^{-\lambda};\ w \in D\}$ with respect to the inner product $\langle \cdot, \cdot \rangle_\lambda$ determined by

$$\langle h(\cdot, w)^{-\lambda}, h(\cdot, z)^{-\lambda} \rangle_\lambda = h(z,w)^{-\lambda}, \quad z, w \in D.$$

Since $h(z,w)^{-\lambda}$ is continuous in $D \times D$, it is the reproducing kernel of $\mathcal{H}_\lambda$. The transformation rule (1.8) implies that $\langle \cdot, \cdot \rangle_\lambda$ is K-invariant, namely $\langle f \circ k, g \circ k \rangle_\lambda = \langle f, g \rangle_\lambda$ for all $f, g \in \mathcal{H}_\lambda$ and $k \in K$. Thus, by Theorems 1.1 and 1.2, for every $f, g \in \mathcal{H}_\lambda$ with Peter-Weyl expansions $f = \sum_{\mathbf{m}\geq 0} f_{\mathbf{m}}$, $g = \sum_{\mathbf{m}\geq 0} g_{\mathbf{m}}$, we have

$$\langle f, g \rangle_\lambda = \sum_{\mathbf{m}\geq 0} \frac{\langle f_{\mathbf{m}}, g_{\mathbf{m}} \rangle_{\mathcal{F}}}{(\lambda)_{\mathbf{m}}}. \tag{1.17}$$

This formula defines $\lambda \mapsto \langle f, g \rangle_\lambda$ as a meromorphic function in all of $\mathbf{C}$, whose poles are contained in the pole set $\mathbf{P}(D)$ of Γ_Ω, see (1.10) and (1.16). Of course, for $\lambda \in \mathbf{C} \setminus \mathbf{W}(D)$ (1.17) is not an inner product, but merely a sesqui-linear form; it is hermitian precisely when $\lambda \in \mathbf{R}$.

Using (1.16) and (1.17) one obtains a complete description of the Wallach set $\mathbf{W}(D)$ and the Hilbert spaces $\mathcal{H}_\lambda$ for $\lambda \in \mathbf{W}(D)$.

THEOREM 1.3 (I) *The Wallach set is given by* $\mathbf{W}(D) = \mathbf{W}_d(D) \cup \mathbf{W}_c(D)$ *where* $\mathbf{W}_d(D) := \{\lambda_j = (j-1)\frac{a}{2}; 1 \leq j \leq r\}$ *is the discrete part, and* $\mathbf{W}_c(D) := (\lambda_r, \infty)$ *is the continuous part.*

(II) *For* $\lambda \in \mathbf{W}_c(D)$ *the polynomials are dense in* $\mathcal{H}_\lambda$ *and* $\mathcal{H}_\lambda = \sum_{\mathbf{m}\geq 0} \oplus P_{\mathbf{m}}$ *as in Theorem 1.1;*

(III) *For* $1 \leq j \leq r$*, let* $S_0(\lambda_j) := \{\mathbf{m} \geq 0; m_j = m_{j+1} = \ldots = m_r = 0\}$*. Then* $\mathcal{H}_{\lambda_j} = \sum_{\mathbf{m}\in S_0(\lambda_j)} P_{\mathbf{m}}$ *and*

$$h(z,w)^{-\lambda_j} = \sum_{\mathbf{m}\in S_0(\lambda_j)} (\lambda_j)_{\mathbf{m}} K_{\mathbf{m}}(z,w), \quad z, w \in D.$$

For $\lambda \in \mathbf{C}$, $\varphi \in G$ and a functions f on D, we define

$$U^{(\lambda)}(\varphi)f := (f \circ \varphi) \cdot (J\varphi)^{\frac{\lambda}{p}}$$

Then, $U^{(\lambda)}(id_D) = I$ and for $\varphi, \psi \in G$ we have

$$U^{(\lambda)}(\varphi \circ \psi) = c_\lambda(\varphi, \psi)\, U^{(\lambda)}(\psi)\, U^{(\lambda)}(\varphi),$$

where $c_\lambda(\varphi, \psi)$ is a unimodular scalar which transforms as a cocycle (*projective representation* of G). In particular, $U^{(\lambda)}(\varphi^{-1}) = U^{(\lambda)}(\varphi)^{-1}$.

Using (1.8) we see that

$$J\varphi(z)^{\frac{\lambda}{p}}\, h(\varphi(z), \varphi(w))^{-\lambda}\, \overline{J\varphi(w)}^{\frac{\lambda}{p}} = h(z, w)^{-\lambda}, \quad \forall z, w \in D, \ \forall \varphi \in G.$$

From this it follows that *the hermitian forms* $\langle \cdot, \cdot \rangle_\lambda$ given by (1.17) are $U^{(\lambda)}$-invariant:

$$\langle U^{(\lambda)}(\varphi)f \,,\, U^{(\lambda)}(\varphi)g \,\rangle_\lambda = \langle f, g\rangle_\lambda, \quad \forall f, g \in \mathcal{H}_\lambda, \quad \forall \varphi \in G.$$

In particular, for $\lambda \in \mathbf{W}(D)$ the inner products $\langle \cdot, \cdot \rangle_\lambda$ are $U^{(\lambda)}$-invariant and $U^{(\lambda)}(\varphi)$, $\varphi \in G$, are unitary operators on $\mathcal{H}_\lambda$.

There are other spaces of analytic functions on D which carry $U^{(\lambda)}$-invariant hermitian forms, some of which are non-negative. For any signature $\mathbf{m}$ and $\lambda \in \mathbf{C}$ let $q(\lambda, \mathbf{m}) := \deg_\lambda(\cdot)_\mathbf{m}$ be the multiplicity of λ as a zero of the polynomial $\xi \mapsto (\xi)_\mathbf{m}$. Notice that $0 \le q(\lambda, \mathbf{m}) \le r$ for all λ and $\mathbf{m}$. Let

$$q(\lambda) := \max\{q(\lambda, \mathbf{m}); \mathbf{m} \ge 0\}. \tag{1.18}$$

Let

$$\mathcal{P}^{(\lambda)} := \operatorname{span}\{U^{(\lambda)}(\varphi)f \ ; \ f \text{ polynomial } , \varphi \in G\}$$

For $0 \le j \le q(\lambda)$ set

$$S_j(\lambda) := \{\mathbf{m} \ge 0; q(\lambda, \mathbf{m}) \le j\} \ \ \mathcal{M}_j^{(\lambda)} := \{f \in \mathcal{P}^{(\lambda)}; f = \sum_{\mathbf{m} \in S_j(\lambda)} f_\mathbf{m}, \ \ f_\mathbf{m} \in P_\mathbf{m}\}. \tag{1.19}$$

The following result is established in [FK1], see also [A1], [O].

Theorem 1.4 *Let $\lambda \in \mathbf{C}$ and let $0 \le j \le q(\lambda)$.*

(i) *The spaces $\mathcal{M}_j^{(\lambda)}, 0 \le j \le q(\lambda)$, are $U^{(\lambda)}$-invariant,*

$$\mathcal{M}_0^{(\lambda)} \subset \mathcal{M}_1^{(\lambda)} \subset \mathcal{M}_2^{(\lambda)} \subset \ldots \subset \mathcal{M}_{q(\lambda)}^{(\lambda)} = \mathcal{P}^{(\lambda)}, \tag{1.20}$$

and every non-zero $U^{(\lambda)}$-invariant subspace of $\mathcal{P}^{(\lambda)}$ is one of the spaces $\mathcal{M}_j^{(\lambda)}$, $0 \le j \le q(\lambda)$.

(ii) *The quotients $\mathcal{M}_j^{(\lambda)}/\mathcal{M}_{j-1}^{(\lambda)}$, $1 \le j \le q(\lambda)$, are $U^{(\lambda)}$-irreducible.*

(III) *The sesqui-linear forms* $\langle\cdot,\cdot\rangle_{\lambda,j}$ *on* $\mathcal{M}_j^{(\lambda)}$, $1 \le j \le q(\lambda)$, *defined for* $f, g \in \mathcal{M}_j^{(\lambda)}$ *by*

$$\langle f, g\rangle_{\lambda,j} := \lim_{\xi\to\lambda} (\xi-\lambda)^j \langle f, g\rangle_\xi$$

are $U^{(\lambda)}$*-invariant and* $\{f \in \mathcal{M}_j^{(\lambda)}; \langle f, g\rangle_{\lambda,j} = 0, \forall g \in \mathcal{M}_j^{(\lambda)}\} = \mathcal{M}_{j-1}^{(\lambda)}$.

(IV) *For* $f, g \in \mathcal{M}_j^{(\lambda)}$ *with Peter-Weyl expansions* $f = \sum_{\mathbf{m}} f_{\mathbf{m}}$ *and* $g = \sum_{\mathbf{m}} g_{\mathbf{m}}$, *we have*

$$\langle f, g\rangle_{\lambda,j} = \sum_{\mathbf{m}\in S_j(\lambda)\setminus S_{j-1}(\lambda)} \frac{\langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}}}{(\lambda)_{\mathbf{m},j}}$$

where

$$(\lambda)_{\mathbf{m},j} := \lim_{\xi\to\lambda} \frac{(\xi)_{\mathbf{m}}}{(\xi-\lambda)^j} = \frac{1}{j!}\left(\frac{d}{d\xi}\right)^j (\xi)_{\mathbf{m}}\,{}_{|\xi=\lambda}. \tag{1.21}$$

(V) *The forms* $\langle\cdot,\cdot\rangle_{\lambda,j}$ *are hermitian if and only if* $\lambda \in \mathbf{R}$.

(VI) *The quotient* $\mathcal{M}_j^{(\lambda)}/\mathcal{M}_{j-1}^{(\lambda)}$ *is unitarizable (namely,* $\langle\cdot,\cdot\rangle_{\lambda,j}$ *is either positive definite or negative definite on* $\mathcal{M}_j^{(\lambda)}/\mathcal{M}_{j-1}^{(\lambda)}$*) if and only if either:* $\lambda \in \mathbf{W}(D)$ *and* $j = 0$, *or:* $\lambda \in \mathbf{P}(D)$, $j = q(\lambda)$, *and* $\lambda_r - \lambda \in \mathbf{N}$.

The sequence (1.20) is called the *composition series* of $\mathcal{P}^{(\lambda)}$.

DEFINITION 1.1 $\mathcal{H}_{\lambda,j} = \mathcal{H}_{\lambda,j}(D)$ *is the completion of* $\mathcal{M}_j^{(\lambda)}/\mathcal{M}_{j-1}^{(\lambda)}$ *with respect to* $\langle\cdot,\cdot\rangle_{\lambda,j}$.

Observe that $\mathcal{H}_{\lambda,0} = \mathcal{H}_\lambda$ for $\lambda \in \mathbf{W}(D)$. Also, $q(\lambda) > 0$ if and only if $\lambda \in \mathbf{P}(D)$.

The *main objective* of this work is to provide natural integral formulas for the $U^{(\lambda)}$-invariant hermitian forms $\langle\cdot,\cdot\rangle_{\lambda,j}$, with special emphasis on the case where the forms are definite, namely the case where $\mathcal{H}_{\lambda,j}$ is a $U^{(\lambda)}$-invariant Hilbert space. These integral formulas provide a characterization of the membership in the spaces $\mathcal{H}_{\lambda,j}$ in terms of finiteness of some weighted L^2-norms of the functions or of some of their derivatives. We discuss now some examples which motivate our study.

The weighted Bergman spaces: It is known [FK1] that for $\lambda \in \mathbf{R}$ the integral $c(\lambda)^{-1} := \int_D h(z,z)^{\lambda-p} dm(z)$ is finite if and only if $\lambda > p-1$, and in this case

$$c(\lambda) = \frac{\Gamma_\Omega(\lambda)}{\pi^d\,\Gamma_\Omega(\lambda - \frac{d}{r})}. \tag{1.22}$$

For $\lambda > p-1$ we consider the probability measure

$$d\mu_\lambda(z) := c(\lambda)\, h(z,z)^{\lambda-p}\, dm(z) \tag{1.23}$$

on D. The *weighted Bergman space* $L_a^2(D, \mu_\lambda)$ consists of all analytic functions in $L^2(D, \mu_\lambda)$. Using (1.8) one obtains the transformation rule of μ_λ under composition with $\varphi \in G$:

$$d\mu_\lambda(\varphi(z)) = |J\varphi(z)|^{\frac{2\lambda}{p}}\, d\mu_\lambda(z).$$

(The same argument yields the invariance of the measure $d\mu_0(z) := h(z,z)^{-p} dm(z)$). From this it follows that the operators $U^{(\lambda)}(\varphi)$ are isometries of $L^2(D, \mu_\lambda)$ which leave $L^2_a(D, \mu_\lambda)$ invariant. It is easy to verify that point evaluations are continuous linear functionals on $L^2_a(D, \mu_\lambda)$ and that the reproducing kernel of $L^2_a(D, \mu_\lambda)$ is $h(z,w)^{-\lambda}$. (For $w = 0$ this is trivial, and the general case follows by invariance.) It follows that $\mathcal{H}_\lambda = L^2_a(D, \mu_\lambda)$.

The Hardy space: The *Shilov boundary* S of a general Cartan domain D is the set of all maximal tripotents in Z. S is invariant and irreducible under both of G and K. Let σ be the unique K-invariant probability measure on S, defined via

$$\int_S f(\xi)\, d\sigma(\xi) := \int_K f(k(e))\, dk.$$

The *Hardy space* $H^2(S)$ is the space of all analytic functions f on D for which

$$\|f\|^2_{H^2(S)} := \lim_{\rho \to 1-} \int_S |f(\rho\xi)|^2\, d\sigma(\xi)$$

is finite. The polynomials are dense in $H^2(S)$ and every $f \in H^2(S)$ has radial limits $\tilde{f}(\xi) := \lim_{\rho\to 1-} f(\rho\xi)$ at σ-almost every $\xi \in S$. Moreover, for $f \in H^2(S)$, $\|f\|_{H^2(S)} = \|\tilde{f}\|_{L^2(S,\sigma)}$. This identifies $H^2(S)$ as the closed subspace of $L^2(S,\sigma)$ consisting of those functions $f \in L^2(S,\sigma)$ which extend analytically to D by means of the Poisson integral. Again, the point evaluations $f \mapsto f(z)$, $z \in D$, are continuous linear functionals on $H^2(S)$. The corresponding reproducing kernel is called the *Szegö kernel* and is given (as a function on S) by $\mathcal{S}_z(\xi) = \mathcal{S}(\xi, z) := h(\xi, z)^{-d/r}$. See [Hu], [FK1]. This non-trivial fact implies that $\mathcal{H}_{d/r} = H^2(S)$. The transformation rule of the measure σ under the automorphisms $\varphi \in G$ is

$$d\sigma(\varphi(\xi)) = |J\varphi(\xi)|\, d\sigma(\xi).$$

Hence, $U^{(d/r)}(\varphi) f = (f \circ \varphi)\,(J\varphi)^{1/2}$, $\varphi \in G$, are isometries of $L^2(S,\sigma)$ which leave $H^2(S)$ invariant.

The Dirichlet space: The classical *Dirichlet space* B_2 consists of those analytic functions f on the open unit disk $\mathbf{D} \subset \mathbf{C}$ for which the *Dirichlet integral*

$$\|f\|^2_{B_2} := \int_{\mathbf{D}} |f'(z)|^2\, dA(z) \tag{1.24}$$

is finite. Here $dA(z) := \frac{1}{\pi} dx\, dy$. Clearly, B_2 is a Hilbert space modulo constant functions, and $\|f \circ \varphi\|_{B_2} = \|f\|_{B_2}$ for every $f \in B_2$ and $\varphi \in Aut(\mathbf{D})$. Thus, B_2 is $U^{(0)}$-invariant. The composition series corresponding to $\lambda = \lambda_1 = 0$ is $\mathbf{C}1 = M_0^{(0)} \subset M_1^{(0)} = \mathcal{P}^{(0)}$. Hence $B_2 = \mathcal{H}_{0,1}(\mathbf{D})$. The inner product in B_2 can be computed also via integration on the boundary $\mathbf{T} := \partial\mathbf{D}$ (which coincides with the Shilov boundary in this simple case):

$$\langle f, g\rangle_{B_2} = \frac{1}{2\pi} \int_{\mathbf{T}} \xi f'(\xi)\, \overline{g(\xi)}\, |d\xi|. \tag{1.25}$$

Motivated by this example we call the spaces $\mathcal{H}_{0,q(0)}$ for a general Cartan domain D the *(generalized) Dirichlet space* of D. The paper [A2] provides integral formulas generalizing (1.24) and (1.25) for the norms in $\mathcal{H}_{\lambda,q(\lambda)}$ for $\lambda \in \mathbf{W}_d(D)$, in the context of a Cartan domain of tube type (in [A1] these formulas are extended to all $\lambda \in \mathbf{P}(D)$). Formula (1.24) says that $f \in B_2 = \mathcal{H}_{0,1}$ if and only if $f' \in \mathcal{H}_2$. Namely, differentiation "shifts" the space corresponding to $\lambda = 0$ to the one corresponding to $\lambda = 2$. This shifting technique is developed in [Y3] in order to get integral formulas for the inner products in certain spaces $\mathcal{H}_\lambda$ with $\lambda \leq p-1$. The general idea is to obtain such integral formulas via "partial integration in the radial directions", see [Ri], [Go], and [FK2], Chapter VII. (For the open unit ball of $\mathbf{C}^d$, the simplest (i.e. rank-one) non-tube Cartan domain, cf. [A3], [Pel]).

Finally, we describe the *relationship between the invariant inner product and the Riesz distribution.* The Riesz distribution was introduced in [Ri] for the Lorentz cone, i.e. the symmetric cone associated with the Cartan domain of type IV (the "Lie ball"). It was studied in [Go] for the cone of symmetric, positive definite real matrices (associated with the Cartan domain of type III) and for a general symmetric cone in [FK2], chapter VII. Let Ω be the symmetric cone associated with the Cartan domain of tube type D. For $\alpha \in \mathbf{C}$ with $\Re\alpha > (r-1)\frac{a}{2}$ let R_α be the linear functional on the Schwartz space $S(X)$ of X defined via

$$R_\alpha(f) := \frac{1}{\Gamma_\Omega(\alpha)} \int_\Omega f(x) N(x)^{\alpha - \frac{d}{r}}\, dx.$$

Then R_α is a tempered distribution satisfying $\partial_N R_\alpha = R_{\alpha-1}$, $R_\alpha \star R_\beta = R_{\alpha+\beta}$, $R_0 = \delta$, i.e. R_1 is the fundamental solution for the "wave operator" $\partial_N := N(\frac{\partial}{\partial x})$. These formulas permit analytic continuation of $\alpha \mapsto R_\alpha$ to an entire meromorphic function. It is very interesting to find the explicit description of the action of R_α for general α, but this is still open. What is known is that *the Riesz distribution R_α is represented by a positive measure if and only if $\alpha \in W(D)$.*

Writing the inner products $\langle\cdot,\cdot\rangle_\lambda$ in conical polar coordinates (1.11), we get for $\lambda > p-1$

$$\langle f,g\rangle_\lambda = \frac{\Gamma_\Omega(\lambda)}{\Gamma_\Omega(\frac{d}{r})\,\Gamma_\Omega(\lambda - \frac{d}{r})} \int_{\Omega\cap(e-\Omega)} (f\overline{g})\check{}(x)\; N(e-x)^{\lambda-p}\, dx, \quad \forall f,g \in \mathcal{H}_\lambda(D),$$

where $(f\bar{g})\check{}(x) := \int_K f(k(x^{\frac{1}{2}}))\,\overline{g(k(x^{\frac{1}{2}}))}\, dk$. Thus

$$\langle f,g\rangle_\lambda = \frac{\Gamma_\Omega(\lambda)}{\Gamma_\Omega(\frac{d}{r})} \left(R_{\lambda-\frac{d}{r}} \star (f\bar{g})\check{}\right)(e),$$

where the convolution of functions u and v on Ω is

$$(u \star v)(x) := \int_{\Omega\cap(x-\Omega)} u(y)\, v(x-y)\, dy.$$

Also, the inner product $\langle\cdot,\cdot\rangle_\lambda$, $\lambda > p-1$, in the context of the tube domain $T(\Omega) := X + i\Omega$ (holomorphically equivalent to D) is

$$\langle f,g\rangle_\lambda := c(\lambda) \int_\Omega \left(\int_X f(x+iy)\,\overline{g(x+iy)}\, dx \right) N(2y)^{\lambda-p}\, dy.$$

See section 6 for the details. Thus

$$\langle f, g\rangle_\lambda = \pi^{-d}\, 2^{\lambda-p}\, \Gamma_\Omega(\lambda)\ R_{\lambda-\frac{d}{r}}\left((f\,\bar{g})^\flat\right),$$

where $(f\,\bar{g})^\flat)(y) := \int_X f(x+iy)\,\overline{g(x+iy)}\,dx,\quad y \in \Omega$.

In view of these formulas the problem of obtaining an explicit description of the analytic continuation of the maps $\lambda \mapsto \langle f, g\rangle_\lambda$ is equivalent to the problem of determining the analytic continuation of the maps $\lambda \mapsto R_{\lambda-\frac{d}{r}}(u)$.

2 $G(\Omega)$-invariant differential operators

Let Ω be the symmetric cone associated with the Cartan domain of tube type D, i.e. the interior of the cone of squares in the Euclidean Jordan algebra X. In this section we study $G(\Omega)$-invariant differential operators that will be used later for the invariant inner products. The ring $\mathrm{Diff}(\Omega)^{G(\Omega)}$ of $G(\Omega)$-invariant differential operators is a (commutative) polynomial ring $\mathbf{C}[X_1, X_2, \ldots, X_r]$, [He], [FK2]. By [FK2], Proposition IX.1.1, Ω is a set of uniqueness for analytic functions on Z (namely, if an analytic function on Z vanishes identically on Ω, it vanishes identically on Z). Similarly, $\Omega \cap D = \Omega \cap (e - \Omega)$ is a set of uniqueness for analytic functions on D. Thus, if f, g and q are polynomials on Z so that $\partial_f(g)(x) = f(\frac{d}{dx})g(x) = q(x)$ for every $x \in \Omega$, then $\partial_f(g)(z) = f(\frac{\partial}{\partial z})g(z) = q(z)$ for every $z \in Z$. We begin with the following known result [FK2], Proposition VII.1.6.

Lemma 2.1 *For every $\mathbf{s} = (s_1, s_2, \ldots, s_r) \in \mathbf{C}^r$ and $\ell \in \mathbf{N}$, we have*

$$N^\ell(\frac{d}{dx})\, N_{\mathbf{s}}(x) = \mu_{\mathbf{s}}(\ell)\ N_{\mathbf{s}-\ell}(x), \quad \forall x \in \Omega,$$

where

$$\mu_{\mathbf{s}}(\ell) := \frac{(\frac{d}{r})_{\mathbf{s}}}{(\frac{d}{r})_{\mathbf{s}-\ell}} = \frac{\Gamma_\Omega(\mathbf{s}+\frac{d}{r})}{\Gamma_\Omega(\mathbf{s}+\frac{d}{r}-\ell)} = \prod_{j=1}^{r}\prod_{\nu=0}^{\ell-1}(s_j - \nu + (r-j)\frac{a}{2}),$$

and

$$\Gamma_\Omega(\mathbf{s})\, N(\frac{d}{dx})\, N_{\mathbf{s}}(x^{-1}) = (-1)^r\ \Gamma_\Omega(\mathbf{s}+1)\ N_{\mathbf{s}+1}(x^{-1}).$$

Let N_j^* be the norm polynomial of the JB*-subalgebra $V_j := \sum_{r-j+1\le j\le k\le r} Z_{i,k}$, where $Z_{i,k}$ are the Peirce subspaces of Z associated with the fixed frame $\{e_j\}_{j=1}^r$. For every $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbf{C}^r$ let

$$N_{\mathbf{s}}^*(x) := N_1^*(x)^{s_1-s_2}\, N_2^*(x)^{s_2-s_3} \ldots N_r^*(x)^{s_r}, \quad x \in \Omega,$$

and

$$\mathbf{s}^* := (s_r, s_{r-1}, s_{r-2}, \ldots, s_1).$$

Then we have $N_{\mathbf{s}}(x^{-1}) = N^*_{-\mathbf{s}^*}(x)$ for $x \in \Omega$, [FK2],Proposition VII.1.5.

Definition 2.1 *For $\ell \in \mathbf{N}$ and $\lambda \in \mathbf{C}$ let $D_\ell(\lambda)$ be the operator on $C^\infty(\Omega)$ defined by*

$$D_\ell(\lambda) = N^{\frac{d}{r}-\lambda}(x)\, N^\ell(\frac{d}{dx})\, N^{\ell+\lambda-\frac{d}{r}}(x). \tag{2.1}$$

In the special case of the Cartan domain of type II the operators $D_1(\lambda)$ have been considered by Selberg (see [T], p.208). The operators $D_\ell(\lambda)$ were studied in full generality in [Y3], see also [FK2], Chapter XIV. Notice that by Lemma 2.1 we have

$$D_\ell(\lambda)\, N_{\mathbf{s}} = \frac{\Gamma_\Omega(\mathbf{s}+\lambda+\ell)}{\Gamma_\Omega(\mathbf{s}+\lambda)}\, N_{\mathbf{s}}. \tag{2.2}$$

In section 4 below we will extend $D_\ell(\lambda)$ to a polynomial differential operator on Z, i.e. $D_\ell(\lambda) = Q_{\ell,\lambda}(z, \frac{\partial}{\partial z})$ for some polynomial $Q_{\ell,\lambda}$.

LEMMA 2.2 *The operator $D_\ell(\lambda)$ is K-invariant, i.e.*

$$D_\ell(\lambda)(f\circ k) = (D_\ell(\lambda)f)\circ k \ \ \forall f \in C^\infty(\Omega), \ \ \forall k \in K.$$

PROOF: We have $N(kz) = \chi(k)N(z)$ for every $z \in Z$. Since the operator $\partial_N = N(\frac{\partial}{\partial z})$ is the adjoint of the operator of multiplication by N with respect to the inner product $\langle\cdot,\cdot\rangle_{\mathcal{F}}$, K-invariance of $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ implies $\partial_N(f\circ k) = \chi(k)(\partial_N f)\circ k$. It follows that

$$\begin{aligned}
D_\ell(\lambda)(f(kz)) &= \overline{\chi(k)}^{\ell+\lambda-\frac{d}{r}}\, N(z)^{\frac{d}{r}-\lambda}\, N^\ell(\frac{\partial}{\partial z})\left(N^{\ell+\lambda-\frac{d}{r}}(kz)f(kz)\right)\\
&= \overline{\chi(k)}^{\ell+\lambda-\frac{d}{r}}\, N(z)^{\frac{d}{r}-\lambda}\, \chi(k)^\ell\, \left(N^\ell(\frac{\partial}{\partial z})(N^{\ell+\lambda-\frac{d}{r}} f)\right)(kz)\\
&= N^{\frac{d}{r}-\lambda}(kz)\, \left(N^\ell(\frac{\partial}{\partial z})(N^{\ell+\lambda-\frac{d}{r}} f)\right)(kz) = (D_\ell(\lambda)f)(kz). \qquad \blacksquare
\end{aligned}$$

Using (2.2) and the fact that $\Omega\cap D = \Omega\cap(e-\Omega)$ is a set of uniqueness for analytic functions on D, we obtain the following result.

COROLLARY 2.1 *The spaces $P_{\mathbf{m}}$ are eigenspaces of $D_\ell(\lambda)$ with eigenvalues*

$$\mu_{\ell,\mathbf{m}}(\lambda) := \frac{\Gamma_\Omega(\mathbf{m}+\lambda+\ell)}{\Gamma_\Omega(\mathbf{m}+\lambda)}. \tag{2.3}$$

Thus for every analytic function f on D with Peter-Weyl expansion $f = \sum_{\mathbf{m}\geq 0} f_{\mathbf{m}}$,

$$D_\ell(\lambda)f = \sum_{\mathbf{m}\geq 0} \frac{\Gamma_\Omega(\mathbf{m}+\lambda+\ell)}{\Gamma_\Omega(\mathbf{m}+\lambda)}\, f_{\mathbf{m}} = (\lambda)_{(\ell,\ell,\dots,\ell)} \sum_{\mathbf{m}\geq 0} \frac{(\lambda+\ell)_{\mathbf{m}}}{(\lambda)_{\mathbf{m}}}\, f_{\mathbf{m}}. \tag{2.4}$$

Indeed, for every signature $\mathbf{m}$ and every $k\in K$,

$$D_\ell(\lambda)(N_{\mathbf{m}}\circ k) = (D_\ell(\lambda)N_{\mathbf{m}})\circ k = \frac{\Gamma_\Omega(\mathbf{m}+\lambda+\ell)}{\Gamma_\Omega(\mathbf{m}+\lambda)}\, N_{\mathbf{m}}\circ k.$$

Since $P_{\mathbf{m}} = span\{N_{\mathbf{m}}\circ k; k\in K\}$, (2.4) follows from the continuity of $D_\ell(\lambda)$ with respect to the topology of compact convergence on D.

COROLLARY 2.2 *Let $\lambda \in \mathbf{C}\setminus \mathbf{P}(D)$, $\ell\in\mathbf{N}$, and $w\in D$. Then*

$$D_\ell(\lambda)\, h(\cdot,w)^{-\lambda} = (\lambda)_{(\ell,\ell,\dots,\ell)}\, h(\cdot,w)^{-(\lambda+\ell)}. \tag{2.5}$$

Proof: Using (1.16) and Corollary 2.2, we get

$$\begin{aligned} D_\ell(\lambda)\, h(\cdot, w)^{-\lambda} &= \sum_{\mathbf{m}\geq 0} (\lambda)_{\mathbf{m}}\, D_\ell(\lambda)\, K_{\mathbf{m}}(\cdot, w) \\ &= (\lambda)_{(\ell,\ldots,\ell)} \sum_{\mathbf{m}\geq 0} (\lambda)_{\mathbf{m}} \frac{(\lambda+\ell)_{\mathbf{m}}}{(\lambda)_{\mathbf{m}}}\, K_{\mathbf{m}}(\cdot, w) \\ &= (\lambda)_{(\ell,\ldots,\ell)} \sum_{\mathbf{m}\geq 0} (\lambda+\ell)_{\mathbf{m}}\, K_{\mathbf{m}}(\cdot, w) = (\lambda)_{(\ell,\ldots,\ell)}\, h(\cdot, w)^{-(\lambda+\ell)}. \quad \blacksquare \end{aligned}$$

Notice that the assumption that λ is not in $\mathbf{P}(D)$ is used in the above proof to ensure that $(\lambda)_{\mathbf{m}} \neq 0$ for every $\mathbf{m} \geq 0$. This is due to the fact that the zero set of the polynomial $(\lambda)_{\mathbf{m}}$ is

$$Z((\cdot)_{\mathbf{m}}) = \cup_{j=1}^{r} \{\lambda_j - k; k = 0, 1, \ldots, m_j - 1\}, \tag{2.6}$$

while $\mathbf{P}(D) = \cup_{j=1}^{r} (\lambda_j - \mathbf{N}) = \cup_{\mathbf{m}\geq 0} Z((\cdot)_{\mathbf{m}})$. Similarly, for each $\mathbf{m} \geq 0$ the zero set of the polynomial defined by (2.3) is given by

$$Z(\mu_{\ell,\mathbf{m}}(\cdot)) = \cup_{j=1}^{r} \{\lambda_j - k;\ m_j \leq k \leq m_j + \ell - 1\}. \tag{2.7}$$

The multiplicities of the zeros are equal to the number of their appearances on the right hand side of (2.7).

Corollary 2.3 *Let $\lambda \in \mathbf{C}$, $\ell \in \mathbf{N}$ be so that $\{\mathbf{m} \geq 0; (\lambda)_{\mathbf{m}} = 0\} \subseteq \{\mathbf{m} \geq 0; (\lambda + \ell)_{\mathbf{m}} = 0\}$. Then (2.5) holds.*

Proof: Notice first that $(\lambda)_{(\ell,\ell,\ldots,\ell)}(\lambda+\ell)_{\mathbf{m}} = (\lambda)_{\mathbf{m}+\ell}$ for all $\lambda \in \mathbf{C}$, $\ell \in \mathbf{N}$, and $\mathbf{m} \geq 0$. Hence, using the fact that $\{\mathbf{m}; (\lambda+\ell)_{\mathbf{m}} \neq 0\} \subseteq \{\mathbf{m}; (\lambda)_{\mathbf{m}} \neq 0\}$, we get for every $w \in D$

$$\begin{aligned} D_\ell(\lambda) h(\cdot, w)^{-\lambda} &= D_\ell(\lambda) \sum_{(\lambda)_{\mathbf{m}}\neq 0} (\lambda)_{\mathbf{m}} K_{\mathbf{m}}(\cdot, w) \\ &= (\lambda)_{(\ell,\ell,\ldots,\ell)} \sum_{(\lambda)_{\mathbf{m}}\neq 0} (\lambda+\ell)_{\mathbf{m}} K_{\mathbf{m}}(\cdot, w) \\ &= (\lambda)_{(\ell,\ldots,\ell)} \sum_{(\lambda+\ell)_{\mathbf{m}}\neq 0} (\lambda+\ell)_{\mathbf{m}} K_{\mathbf{m}}(\cdot, w) \\ &= (\lambda)_{(\ell,\ldots,\ell)} h(\cdot, w)^{-(\lambda+\ell)}. \quad \blacksquare \end{aligned}$$

For $\lambda \in \mathbf{P}(D)$ let $q = q(\lambda)$ be as in (1.18), and for $0 \leq j \leq q$ consider $S_j(\lambda)$ and $\mathcal{M}_j^{(\lambda)}$ as in (1.19).

Lemma 2.3 *Let λ, and $q = q(\lambda)$ be as above, and choose an integer ℓ so that $\lambda + \ell \geq \frac{d}{r} = \lambda_r + 1$. Then*

(i) *$deg_\lambda((\cdot)_{(\ell,\ell,\ldots,\ell)}) = q$.*

(II) *For every* $j = 0, 1, 2, \ldots, q$ *and every* $\mathbf{m} \in S_j(\lambda) \setminus S_{j-1}(\lambda)$, $deg_\lambda(\mu_{\ell,\mathbf{m}}) = q - j$.

(III) *If* $0 \leq j \leq q$ *and* $\mathbf{m} \in S_{j-1}(\lambda)$, *then* $deg_\lambda(\mu_{\ell,\mathbf{m}}) \geq q - j + 1$.

PROOF: Using (2.6) it is clear that

$$q(\lambda, \mathbf{m}) = q \quad \Leftrightarrow \quad \lambda_j - m_j + 1 \leq \lambda \ \ \forall j \quad \Leftrightarrow \quad \lambda_r - m_r + 1 \leq \lambda.$$

Since $\lambda_r + 1 \geq \lambda + \ell$, we see that $\mathbf{m} = (\ell, \ell, \ldots, \ell)$ satisfies the above condition, namely $deg_\lambda((\cdot)_{(\ell,\ldots,\ell)}) = q(\lambda, (\ell, \ldots, \ell)) = q$. This establishes (i). Next, $\mathbf{m} \in S_j^{(\lambda)} \setminus S_{j-1}^{(\lambda)}$ is equivalent to $q(\lambda, \mathbf{m}) = j$. By the argument given above, $q(\lambda, \mathbf{m} + \ell) = q$. Since $deg_\lambda(f/g) = deg_\lambda(f) - deg_\lambda(g)$, we get

$$deg_\lambda(\mu_{\ell,\mathbf{m}}) = deg_\lambda\left(\frac{(\cdot)_{\mathbf{m}+\ell}}{(\cdot)_{\mathbf{m}}}\right) =$$

$$= deg_\lambda((\cdot)_{\mathbf{m}+\ell}) - deg_\lambda((\cdot)_{\mathbf{m}}) = q(\lambda, \mathbf{m} + \ell) - q(\lambda, \mathbf{m}) = q - j.$$

This yields (ii). Finally, (iii) follows by similar computations. ∎

Let $\lambda \in \mathbf{P}(D)$, $\ell \in \mathbf{N}$, and $q = q(\lambda)$ as above. For every $\mathbf{m} \geq 0$ and $\nu \in \mathbf{N}$ we define

$$\mu^\nu_{\ell,\mathbf{m}}(\lambda) := \frac{1}{\nu!}(\frac{\partial}{\partial \xi})^\nu \mu_{\ell,\mathbf{m}}(\xi)_{|\xi=\lambda}.$$

Using Lemma 2.3 (ii), we have

COROLLARY 2.4 (I) *If* $\mathbf{m} \in S_j(\lambda) \setminus S_{j-1}(\lambda)$ *then*

$$\mu^{q-j}_{\ell,\mathbf{m}}(\lambda) = \prod_{i=1}^r {\prod}'^{\,m_i+\ell-1}_{k=m_i} (\lambda + k - \lambda_i),$$

where the product ${\prod}'^{\,m_j+\ell-1}_{k=m_j}$ *ranges over all non-zero terms. In particular,* $\mu^{q-j}_{\ell,\mathbf{m}}(\lambda) \neq 0$.

(II) *If* $\mathbf{m} \in S_{j-1}(\lambda)$ *then* $\mu^{q-j}_{\ell,\mathbf{m}}(\lambda) = 0$.

DEFINITION 2.2 *For* $\lambda \in \mathbf{C}$ *and* $\nu, \ell \in \mathbf{N}$ *let* $D^\nu_\ell(\lambda)$ *be the operator on* $C^\infty(D)$ *defined by*

$$D^\nu_\ell(\lambda)f := \frac{1}{\nu!}(\frac{\partial}{\partial \xi})^\nu (D_\ell(\xi)f)_{|\xi=\lambda}. \tag{2.8}$$

Notice that if $f = \sum_{\mathbf{m}\geq 0} f_{\mathbf{m}}$ is analytic in D, then $D^\nu_\ell(\lambda)f := \sum_{\mathbf{m}\geq 0} \mu^\nu_{\ell,\mathbf{m}}(\lambda)\, f_{\mathbf{m}}$.

By [FK2], Chapter VI the group $G(\Omega)$ admits an Iwasawa decomposition $G(\Omega) = NAL$, where L is the group defined via (1.6), and NA is a maximal solvable subgroup of $G(\Omega)$ (called the *triangular subgroup* with respect to the frame $\{e_i\}_{i=1}^r$) which acts simply transitively on Ω and for which all the conical functions $N_{\mathbf{s}}$, $\mathbf{s} \in \mathbf{C}^r$, are eigenfunctions:

$$N_{\mathbf{s}}(\tau(x)) = N_{\mathbf{s}}(\tau(e))\, N_{\mathbf{s}}(x), \quad \forall \tau \in NA, \quad \forall x \in \Omega. \tag{2.9}$$

Lemma 2.4 *The operators* $D_\ell(\lambda)$ *are* $G(\Omega)$*-invariant, i.e.* $D_\ell(\lambda)(f\circ\varphi) = (D_\ell(\lambda)f)\circ\varphi,\ \forall f\in C^\infty(\Omega),\ \forall\varphi\in G(\Omega)$.

Proof: By the L-invariance of $D_\ell(\lambda)$ (see Lemma 2.2) it is enough to verify the NA-invariance of $D_\ell(\lambda)$ for functions f of the form $f = N_{\mathbf{s}}\circ\ell$ for some $\mathbf{s}\in\mathbf{C}^r$ and $\ell\in L$. Let $\tau\in NA$, and decompose $\ell\circ\tau$ uniquely as $\ell\circ\tau = \tau'\circ\ell'$ with $\tau'\in NA$ and $\ell'\in L$. Then, using (2.2) and (2.9), we get

$$\begin{aligned}
D_\ell(\lambda)(f\circ\tau) &= D_\ell(\lambda)(N_{\mathbf{s}}\circ\ell\circ\tau) = D_\ell(\lambda)(N_{\mathbf{s}}\circ\tau'\circ\ell')\\
&= (D_\ell(\lambda)(N_{\mathbf{s}}\circ\tau'))\circ\ell' = N_{\mathbf{s}}(\tau'(e))(D_\ell(\lambda)N_{\mathbf{s}})\circ\ell'\\
&= N_{\mathbf{s}}(\tau'(e))\,\frac{\Gamma_\Omega(\mathbf{s}+\lambda+\ell)}{\Gamma_\Omega(\mathbf{s}+\lambda)}\,N_{\mathbf{s}}\circ\ell' \;=\; \frac{\Gamma_\Omega(\mathbf{s}+\lambda+\ell)}{\Gamma_\Omega(\mathbf{s}+\lambda)}\,N_{\mathbf{s}}\circ\tau'\circ\ell'\\
&= \frac{\Gamma_\Omega(\mathbf{s}+\lambda+\ell)}{\Gamma_\Omega(\mathbf{s}+\lambda)}\,N_{\mathbf{s}}\circ\ell\circ\tau \;=\; \frac{\Gamma_\Omega(\mathbf{s}+\lambda+\ell)}{\Gamma_\Omega(\mathbf{s}+\lambda)}\,f\circ\tau\\
&= (D_\ell(\lambda)f)\circ\tau.
\end{aligned}$$

∎

Corollary 2.5 *The operators* $D_\ell^\nu(\lambda)$ *are* $G(\Omega)$*-invariant.*

3 Integral formulas via the shifting method

In this section we develop general shifting techniques (introduced in [Y3], for the case of integer shifts). The simplest case where this technique is applied is the case of the Dirichlet space $\mathcal{D} = \mathcal{H}_{0,1}$ over the unit disk $\mathbf{D}$ (see Section 2). For any $\alpha\in\mathbf{C}$ and $\beta\in\mathbf{C}\setminus\mathbf{P}(D)$ we define an operator $S_{\alpha,\beta}$ on $\mathcal{H}(D)$ via

$$S_{\alpha,\beta}\Big(\sum_{\mathbf{m}\geq 0} f_{\mathbf{m}}\Big) := \sum_{\mathbf{m}\geq 0}\frac{(\alpha)_{\mathbf{m}}}{(\beta)_{\mathbf{m}}}\,f_{\mathbf{m}}.$$

Theorem 5 of [A4] and the known estimate

$$\frac{(x)_{\mathbf{m}}}{(y)_{\mathbf{m}}} \approx \prod_{j=1}^r (m_j+1)^{x-y},\quad \forall x,y\in\mathbf{R}$$

(an easy consequence of (1.9) and Stirling's formula) ensures that $S_{\alpha,\beta}$ is continuous on $\mathcal{H}(D)$. For $\beta\in\mathbf{P}(D)$ we define operators $S_{\alpha,\beta,j}$, $0\leq j\leq q(\beta)$, on the space of analytic functions on D of the form $f = \sum_{\mathbf{m}\in S_j(\beta)} f_{\mathbf{m}}$ by

$$S_{\alpha,\beta,j}f := \lim_{\xi\to\beta}(\xi-\beta)^j S_{\alpha,\beta}f = \sum_{\mathbf{m}\in S_j(\beta)\setminus S_{j-1}(\beta)}\frac{(\alpha)_{\mathbf{m}}}{(\beta)_{\mathbf{m},j}}\,f_{\mathbf{m}},$$

where $(\beta)_{\mathbf{m},j}$ are defined by (1.21). Again, $S_{\alpha,\beta,j}$ is continuous in the topology of $\mathcal{H}(D)$. Also, $S_{\alpha,\beta,0} = S_{\alpha,\beta}$.

PROPOSITION 3.1 *Let* $\alpha, \beta > (r-1)\frac{a}{2}$. *Then* $\langle f, g\rangle_\beta = \langle S_{\alpha,\beta} f, g\rangle_\alpha$ *for every* $f, g \in \mathcal{H}_\beta$.

PROOF: By (1.17) the operator $S^{\frac{1}{2}}_{\alpha,\beta} : \mathcal{H}_\beta \to \mathcal{H}_\alpha$ defined by

$$S^{\frac{1}{2}}_{\alpha,\beta}(\sum_{\mathbf{m}\geq 0} f_{\mathbf{m}}) := \sum_{\mathbf{m}\geq 0} \left(\frac{(\alpha)_{\mathbf{m}}}{(\beta)_{\mathbf{m}}}\right)^{\frac{1}{2}} f_{\mathbf{m}}$$

is a surjective isometry, and $\|f\|^2_\beta = \|S^{\frac{1}{2}}_{\alpha,\beta} f\|^2_\alpha = \langle S_{\alpha,\beta} f, f\rangle_\alpha$. Now the result follows by polarization. ∎

In a similar way one proves the following result.

PROPOSITION 3.2 *Let* $\alpha > (r-1)\frac{a}{2}$ *and let* $\beta \in \mathbf{P}(D)$. *Then for every* $0 \leq j \leq q(\beta)$ *and all* $f, g \in \mathcal{H}_{\beta,j}$,

$$\langle f, g\rangle_{\beta,j} = \langle S_{\alpha,\beta,j} f, g\rangle_\alpha. \tag{3.1}$$

The operators $S_{\alpha,\beta,j}$ allow the computation of the invariant hermitian forms $\langle f, g\rangle_{\beta,j}$ by "shifting" the point β to the point α. This is the "*shifting method*". One typically chooses either $\alpha = \frac{d}{r}$ or $\alpha > p-1$, so the forms $\langle f, g\rangle_{\beta,j}$ can be computed by the integral-type inner products of $H^2(D)$ or $L^2_a(D, \mu_\alpha)$. In order for the shifting method to be useful, one has to identify the operators $S_{\alpha,\beta,j}$ as differential or pseudo-differential operators. Essentially, this is our aim in the rest of the paper. Yan's paper [Y3] deals with the case where $\ell := \alpha - \beta$ is a sufficiently large natural number. The following result is a minor generalization of a result of [Y3].

THEOREM 3.1 *Let* $\lambda > \lambda_r = \frac{d}{r} - 1$ *and let* $\ell \in \mathbf{N}$. *Then for all* $f, g \in \mathcal{H}_\lambda$

$$\langle f, g\rangle_\lambda = \alpha(\lambda, \ell)\langle D_\ell(\lambda) f, g\rangle_{\lambda+\ell}, \tag{3.2}$$

where

$$\alpha(\lambda, \ell) = \frac{\Gamma_\Omega(\lambda)}{\Gamma_\Omega(\lambda+\ell)} = \frac{1}{(\lambda)_{(\ell,\ell,\ldots,\ell)}}.$$

We include a short proof for the sake of completeness.

PROOF: Let $f, g \in \mathcal{H}_\lambda$ with expansions $f = \sum_{\mathbf{m}\geq 0} f_{\mathbf{m}}$ and $g = \sum_{\mathbf{m}\geq 0} g_{\mathbf{m}}$ respectively. Then

$$\begin{aligned} \langle D_\ell(\lambda) f, g\rangle_{\lambda+\ell} &= \sum_{\mathbf{m}\geq 0} \frac{\mu_{\ell,\mathbf{m}}(\lambda)}{(\lambda+\ell)_{\mathbf{m}}} \langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}} \\ &= \frac{\Gamma_\Omega(\lambda+\ell)}{\Gamma_\Omega(\lambda)} \sum_{\mathbf{m}\geq 0} \frac{\langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}}}{(\lambda)_{\mathbf{m}}} = \alpha(\lambda,\ell)^{-1}\langle f, g\rangle_\lambda. \end{aligned}$$ ∎

COROLLARY 3.1 *Let* $\lambda > \lambda_r = \frac{d}{r} - 1$, *and* $\ell \in \mathbf{N}$ *be so that* $\lambda + \ell > p - 1$. *Then* $\mathcal{H}_{\lambda+\ell} = L^2_a(D, \mu_{\lambda+\ell})$, *and for every* $f, g \in L^2_a(D, \mu_{\lambda+\ell})$,

$$\langle f, g\rangle_\lambda = \alpha(\lambda, \ell)\, c(\lambda+\ell) \int_D (D_\ell(\lambda) f)(z)\, \overline{g(z)}\, h(z,z)^{\lambda+\ell-p}\, dm(z).$$

Our main result in this section is a generalization of both Theorem 3.1 and the other results of [Y3] to the case of invariant hermitian forms associated with the pole set $\mathbf{P}(D) = \cup_{j=1}^{r}(\lambda_j - \mathbf{N})$. Since $\mathbf{W}(D) \subset \mathbf{P}(D)$, this covers cases not studied in [A1].

Theorem 3.2 *Let $\lambda \in \mathbf{P}(D)$, $\ell \in \mathbf{N}$ and assume that $\lambda + \ell \geq \frac{d}{r} = \lambda_r + 1$. Let $q = q(\lambda)$, $0 \leq j \leq q$, and $\nu = q - j$. Then for all $f, g \in \mathcal{H}_{\lambda,j}$,*

$$\langle f, g\rangle_{\lambda,j} = \gamma \langle D_\ell^\nu(\lambda) f, g\rangle_{\lambda+\ell}, \tag{3.3}$$

where $\gamma = \gamma(\lambda, \ell)$ is the non-zero constant

$$\gamma := \frac{1}{q!}(\frac{\partial}{\partial \xi})^q \left((\xi)_{(\ell,\ell,\ldots,\ell)}\right)_{|\xi=\lambda}. \tag{3.4}$$

In particular, if $\lambda + \ell > p - 1$, then

$$\langle f, g\rangle_{\lambda,j} = \gamma\, c(\lambda+\ell) \int_D (D_\ell^\nu(\lambda) f)(z)\, \overline{g(z)}\, dm(z). \tag{3.5}$$

Moreover, if $\lambda_r - \lambda \in \mathbf{N}$ and ℓ is chosen so that $\lambda + \ell = \frac{d}{r} = \lambda_r + 1$, then

$$\langle f, g\rangle_{\lambda,j} = \gamma \int_S (D_\ell^\nu(\lambda) f)(\xi)\, \overline{g(\xi)}\, d\sigma(\xi). \tag{3.6}$$

We shall also give a new proof of the following known result (see [FK1], Theorem 5.3) and of a part of Theorem 1.4 above, based on our analysis of the structure of zeros of the polynomials $(\cdot)_{\mathbf{m}}$. Recall that $\mathcal{H}_{\lambda,j}$ is said to be *unitarizable* if $\langle\cdot,\cdot\rangle_{\lambda,j}$ is either positive definite or negative definite.

Theorem 3.3 *Let λ, ℓ, q, and j be as in Theorem 3.2. Then $\mathcal{H}_{\lambda,j}$ is unitarizable if and only if either*
(a) $j = q$ and $\lambda_r - \lambda \in \mathbf{N}$, or
(b) $j = 0$ and $\lambda \in \mathbf{W}_d(D) = \{\lambda_j\}_{j=1}^r$.

For the proof of Theorems 3.2 and 3.3 we consider separately the cases $j = 0$, $j = q$, and $1 \leq j \leq q-1$.

Case 1: $\mathbf{j = 0}$. Since $\lambda \in \mathbf{P}(D)$, there is a smallest $k \in \{1, 2, \ldots, r\}$ and a unique $s \in \mathbf{N}$ so that $\lambda = \lambda_k - s$. We claim that $S_0(\lambda) = \{\mathbf{m} \geq 0; m_k \leq s\}$. Indeed, if $\mathbf{m} \geq 0$, then $\prod_{i=1}^{k-1}\prod_{\nu=0}^{m_i-1}(\lambda + \nu - \lambda_i) \neq 0$, by the minimality of k. The term $\prod_{\nu=0}^{m_k-1}(\lambda+\nu-\lambda_k) = \prod_{\nu=0}^{m_k-1}(\nu - s)$ is non-zero if and only if $m_k \leq s$. If $m_k \leq s$ and $k < n \leq r$ then

$$\prod_{\nu=0}^{m_k-1}(\lambda+\nu-\lambda_k) = \prod_{\nu=0}^{m_k-1}((\lambda_k - \lambda_n) + (\nu - s)) \neq 0$$

because $m_n \leq m_k \leq s$. This establishes the claim. Notice that since $\lambda + \ell \geq \lambda_r + 1$, we have $(\lambda+\ell)_{\mathbf{m}} > 0$ for any $\mathbf{m} \geq 0$. Also, $deg_\lambda((\cdot)_{(\ell,\ell,\ldots,\ell)}) = q$ by Lemma 2.3. It follows that for $\mathbf{m} \in S_0(\lambda)$, $deg_\lambda(\mu_{\ell,\mathbf{m}}) = q$, and

$$\begin{aligned}\mu^q_{\ell,\mathbf{m}}(\lambda) &= \frac{1}{q!}(\frac{\partial}{\partial\xi})^q \mu_{\ell,\mathbf{m}}(\xi)_{\ |\xi=\lambda} = \frac{1}{q!}(\frac{\partial}{\partial\xi})^q \left(\frac{(\xi+\ell)_{\mathbf{m}}}{(\xi)_{\mathbf{m}}}(\xi)_{(\ell,\ell,\ldots,\ell)}\right)_{|\xi=\lambda} \\ &= \frac{(\lambda+\ell)_{\mathbf{m}}}{(\lambda)_{\mathbf{m}}}\frac{1}{q!}(\frac{\partial}{\partial\xi})^q(\xi)_{(\ell,\ell,\ldots,\ell)\ |\xi=\lambda} = \gamma\ \frac{(\lambda+\ell)_{\mathbf{m}}}{(\lambda)_{\mathbf{m}}}.\end{aligned}$$

Hence, for $f, g \in \mathcal{H}_{\lambda,0}$,

$$\begin{aligned}\langle D^q_\ell(\lambda)f,g\rangle_{\lambda+\ell} &= \sum_{\mathbf{m}\in S_0(\lambda)} \mu^q_{\ell,\mathbf{m}}(\lambda)\ \frac{\langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}}}{(\lambda+\ell)_{\mathbf{m}}} \\ &= \gamma \sum_{\mathbf{m}\in S_0(\lambda)} \frac{\langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}}}{(\lambda)_{\mathbf{m}}} = \gamma\langle f,g\rangle_{\lambda,0}.\end{aligned}$$

This proves Theorem 3.2 in case $j = 0$. If $\lambda \in \mathbf{W}_d(D)$, i.e. $\lambda = \lambda_k$ and $s = 0$, then $(\lambda)_{\mathbf{m}} > 0$ for every $\mathbf{m} \in S_0(\lambda)$, namely $0 = m_k = m_{k+1} = \cdots = m_r$. If $\lambda \in \mathbf{P}(D) \setminus \mathbf{W}_d(D)$, then $\lambda = \lambda_k - s$ with $1 \leq s$. In this case $(\lambda)_{\mathbf{m}}$ assumes both positive and negative values as $\mathbf{m}$ ranges over $S_0(\lambda)$. Indeed, if $\mathbf{m}$ and $\mathbf{n}$ are defined by $m_i = n_i = 0$ for $1 \leq i \leq k-1$ and $k < i \leq r$, and $m_k = 0$, $n_k = s-1$, then $(\lambda)_{\mathbf{m}}$ and $(\lambda)_{\mathbf{n}}$ have different signs. Thus $\langle\cdot,\cdot\rangle_{\lambda,0}$ is not definite (positive or negative), and thus $\mathcal{H}_{\lambda,0}$ is not unitarizable. This proves Theorem 3.3 in case $j = 0$.

CASE 2: $\mathbf{j} = \mathbf{q}$. In this case $\nu = q - j = 0$. Also, Lemma 2.3 yields $deg_\lambda(\mu_{\ell,\mathbf{m}}) = 0$ if $\mathbf{m} \in S_q(\lambda)$ and $deg_\lambda(\mu_{\ell,\mathbf{m}}) \geq 1$ if $\mathbf{m} \in S_{q-1}(\lambda)$. It follows that for $f, g \in \mathcal{H}_{\lambda,q}$,

$$\langle D_\ell(\lambda)f,g\rangle_{\lambda+\ell} = \sum_{\mathbf{m}\in S_q(\lambda)} \mu_{\ell,\mathbf{m}}(\lambda)\frac{\langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}}}{(\lambda+\ell)_{\mathbf{m}}}.$$

Now,

$$\mu_{\ell,\mathbf{m}}(\lambda) = \lim_{\xi\to\lambda}\frac{(\xi+\ell)_{\mathbf{m}}}{(\xi)_{\mathbf{m}}}\ (\xi)_{(\ell,\ell,\ldots,\ell)} = (\lambda+\ell)_{\mathbf{m}}\lim_{\xi\to\lambda}\frac{(\xi)_{(\ell,\ell,\ldots,\ell)}}{(\xi)_{\mathbf{m}}} = \gamma\frac{(\lambda+\ell)_{\mathbf{m}}}{(\lambda)_{\mathbf{m},q}},$$

where γ is the non-zero constant defined in (3.4). It follows that

$$\langle D_\ell(\lambda)f,g\rangle_{\lambda+\ell} = \gamma\sum_{\mathbf{m}\in S_q(\lambda)}\frac{\langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}}}{(\lambda)_{\mathbf{m},q}} = \gamma\langle f,g\rangle_{\lambda,q}.$$

This proves Theorem 3.2 in case $j = q$. To prove Theorem 3.3 in this case, assume first that $\lambda = \lambda_r - s$ for some $s \in \mathbf{N}$. We claim now that

$$S_q(\lambda) \setminus S_{q-1}(\lambda) = \{\mathbf{m} \geq 0; m_r \geq s+1\}. \tag{3.7}$$

Indeed, if $m_r \geq s+1$ then $\prod_{u=0}^{m_r-1}(\lambda + u - \lambda_r) = 0$. If $\lambda \in \lambda_i - \mathbf{N}$, then $\prod_{u=0}^{m_i-1}(\lambda+u-\lambda_r) = 0$ because $m_i \geq m_r \geq s+1$. Thus $deg_\lambda((\cdot)_{\mathbf{m}}) = q$. Conversely, if $deg_\lambda((\cdot)_{\mathbf{m}}) = q$, then in order that $\prod_{u=0}^{m_r-1}(\lambda+u-\lambda_r)=0$ it is necessary that $s \le m_r - 1$. This establishes (3.7).

Next, let $\mathbf{m} \in S_q(\lambda)$, and let $1 \le i \le r$ be so that $\lambda \in \lambda_i - \mathbf{N}$, say $\lambda = \lambda_i - k_i$. Then

$$\lim_{\xi\to\lambda}(\xi-\lambda)^{-1}\prod_{u=0}^{m_i-1}(\xi+u-\lambda_i) = \prod_{u=0}^{k_i-1}(\lambda+u-\lambda_i)\prod_{u=k_i+1}^{m_i-1}(\lambda+u-\lambda_i) = \gamma_{i,\mathbf{m}}\,\beta_i$$

with $\beta_i \neq 0$ and $\gamma_{i,\mathbf{m}} > 0$. If $\lambda \notin \lambda_i - \mathbf{N}$ we let $\beta_i = \prod_{u<\lambda_i-\lambda}(\lambda+u-\lambda_i) \neq 0$ and $\gamma_{i,\mathbf{m}} = \prod_{u>\lambda_i-\lambda}(\lambda+u-\lambda_i) > 0$. Then

$$(\lambda)_{\mathbf{m},q} = \lim_{\xi\to\lambda}\frac{(\xi)_{\mathbf{m}}}{(\xi-\lambda)^q} = \prod_{i=1}^{r}\gamma_{i,\mathbf{m}}\,\beta_i.$$

Hence, all the numbers $\{(\lambda)_{\mathbf{m},q}\}_{\mathbf{m}\in S_q(\lambda)}$ have constant sign (equal to $sgn(\prod_{i=1}^r\beta_i)$), and thus $\mathcal{H}_{\lambda,q}$ is unitarizable. Assume now that $\lambda \notin \lambda_r - \mathbf{N}$. Then, necessarily, the characteristic multiplicity a is odd and $\lambda \in \lambda_{r-1} - \mathbf{N}$. Writing $\lambda = \lambda_{r-1} - s$, $s \in \mathbf{N}$, it is clear by the above arguments that

$$S_q(\lambda) \setminus S_{q-1}(\lambda) = \{\mathbf{m} \ge 0;\ m_{r-1} \ge s+1\}.$$

Let $\mathbf{m} = (s+1, s+1, \ldots, s+1, 1)$ and $\mathbf{n} = (s+1, s+1, \ldots, s+1, 0)$. Then $\mathbf{m}, \mathbf{n} \in S_q(\lambda)$ and $(\lambda)_{\mathbf{m},q} = (\lambda-\lambda_r)(\lambda)_{\mathbf{n},q}$. Thus $(\lambda)_{\mathbf{m},q}$ and $(\lambda)_{\mathbf{n},q}$ have different signs, and so $\mathcal{H}_{\lambda,q}$ is not unitarizable. This proves Theorem 3.3 in case $j = q$.

Case 3: $\mathbf{1 \le j \le q-1}$. Put $\nu = q - j$. As before, $\ell \in \mathbf{N}$ is chosen so that $\lambda + \ell \ge \lambda_r + 1$, and this guarantees that $deg_\lambda((\cdot)_{\mathbf{m}+\ell}) = q$ and $(\lambda+\ell)_{\mathbf{m}} > 0$ for all signatures $\mathbf{m} \ge 0$. Let $f, g \in \mathcal{H}_{\lambda,j}$. Then

$$\langle D_\ell^\nu(\lambda)f, g\rangle_{\lambda+\ell} = \sum_{\mathbf{m}\in S_j(\lambda)}\mu_{\ell,\mathbf{m}}^\nu(\lambda)\,\frac{\langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}}}{(\lambda+\ell)_{\mathbf{m}}}.$$

If $\mathbf{m} \in S_j(\lambda) \setminus S_{j-1}(\lambda)$, then

$$deg_\lambda(\mu_{\ell,\mathbf{m}}) = deg_\lambda\left(\frac{(\cdot)_{\mathbf{m}+\ell}}{(\cdot)_{\mathbf{m}}}\right) = q - j = \nu.$$

Thus,

$$\mu_{\ell,\mathbf{m}}^\nu(\lambda) = \lim_{\xi\to\lambda}\frac{\mu_{\ell,\mathbf{m}}(\xi)}{(\xi-\lambda)^\nu} = \lim_{\xi\to\lambda}\frac{(\xi+\ell)_{\mathbf{m}}(\xi-\lambda)^{-q}(\xi)_{(\ell,\ell,\ldots,\ell)}}{(\xi-\lambda)^{-j}(\xi)_{\mathbf{m}}} = \gamma\,\frac{(\lambda+\ell)_{\mathbf{m}}}{(\lambda)_{\mathbf{m},j}}.$$

If $\mathbf{m} \in S_{j-1}(\lambda)$, then $deg_\lambda(\mu_{\ell,\mathbf{m}}) \ge q - j + 1 = \nu + 1$, and so $\mu_{\ell,\mathbf{m}}^\nu(\lambda) = 0$. Thus

$$\begin{aligned}\langle D_\ell^\nu(\lambda)f, g\rangle_{\lambda+\ell} &= \gamma\sum_{\mathbf{m}\in S_j(\lambda)\setminus S_{j-1}(\lambda)}\frac{(\lambda+\ell)_{\mathbf{m}}}{(\lambda)_{\mathbf{m},j}}\,\frac{\langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}}}{(\lambda+\ell)_{\mathbf{m}}}\\ &= \gamma\sum_{\mathbf{m}\in S_j(\lambda)\setminus S_{j-1}(\lambda)}\frac{\langle f_{\mathbf{m}}, g_{\mathbf{m}}\rangle_{\mathcal{F}}}{(\lambda)_{\mathbf{m},j}} = \gamma\langle f, g\rangle_{\lambda,j}.\end{aligned}$$

This proves Theorem 3.2 in case $1 \leq j \leq q-1$. To prove Theorem 3.3 in this case we need to show that as $\mathbf{m}$ varies in $S_j(\lambda) \setminus S_{j-1}(\lambda)$, $(\lambda)_{\mathbf{m},j}$ assumes both positive and negative values. Notice first that there exists a unique pair (k,s) of integers with $1 \leq k < s \leq r$ so that $\lambda_k - \lambda$ and $\lambda_s - \lambda$ are positive integers and

$$\mathbf{m} \in S_j(\lambda) \setminus S_{j-1}(\lambda) \quad \Longleftrightarrow \quad m_k \geq \lambda_k - \lambda + 1 \ \ and \ \ m_s \leq \lambda_s - \lambda.$$

In fact, $s = k+1$ if the characteristic multiplicity a is even, and $s = k+2$ if a is odd. Next, $\lambda_s - \lambda = \lambda_k - \lambda + (s-k)\frac{a}{2} \geq 1$. Define $\mathbf{m}, \mathbf{n}$ by $m_i = n_i = \lambda_k - \lambda + 1$ if $1 \leq i \leq k$, $m_i = n_i = 0$ if $k+2 \leq i \leq r$, and $m_{k+1} = 0$, $n_{k+1} = 1$. Then $\mathbf{m}, \mathbf{n} \in S_j(\lambda) \setminus S_{j-1}(\lambda)$ and $(\lambda)_{\mathbf{n},j} = (\lambda)_{\mathbf{m},j}(\lambda - \lambda_s)$. Thus $(\lambda)_{\mathbf{n},j}$ and $(\lambda)_{\mathbf{m},j}$ have different signs, and so $\mathcal{H}_{\lambda,j}$ is not unitarizable. This proves Theorem 3.3 in case $1 \leq j \leq q-1$. ∎

A special case of Theorem 3.2 is the following essentially known result.

COROLLARY 3.2 *Let $\lambda \in \mathbf{P}(D)$ be so that $s = s(\lambda) := \frac{d}{r} - \lambda \in \mathbf{N}$. Then*

(I) *$\mathcal{H}_{\lambda,q}$ is unitarizable, and*

$$\langle f, g \rangle_{\lambda,q} = \gamma \int_S N^s(\xi)(\partial_N^s f)(\xi)\, \overline{g(\xi)}\, d\sigma(\xi), \quad \forall f, g \in \mathcal{H}_{\lambda,q}.$$

Thus, an analytic function f on D belongs to $\mathcal{H}_{\lambda,q}$ if and only if $(N^s \partial_N^s)^{1/2} f \in H^2(S)$.

(II) *Moreover, if $\ell \in \mathbf{N}$ is chosen so that $\lambda + \ell > p - 1$, then*

$$\langle f, g \rangle_{\lambda,q} = \gamma' \int_D (D_\ell(\lambda) f)(z)\, \overline{g(z)}\, h(z,z)^{\lambda+\ell-p}\, dm(z), \quad \forall f, g \in \mathcal{H}_{\lambda,q}.$$

Consequently, an analytic function f on D belongs to $\mathcal{H}_{\lambda,q}$ if and only if $(D_\ell(\lambda))^{1/2} f \in L^2_a(D, \mu_{\lambda+\ell})$.

In the last statement $(D_\ell(\lambda))^{1/2}$ is the positive square root of the positive operator $D_\ell(\lambda)$, see Corollary 2.1 Indeed, part (i) follows from Theorem 3.2 with $j = q$, $\nu = q - j = 0$, $\ell = s$ and $D_s(\lambda) = N^s \partial_N^s$. In this case $\mathcal{H}_{\lambda+s} = \mathcal{H}_{\frac{d}{r}}$ is the Hardy space $H^2(S)$ on the Shilov boundary S. Corollary 3.2 (i) for $\lambda \in \mathbf{W}_d(D)$ was proved in [A2]. The proof of part (ii) is similar.

The case where $\lambda \in \mathbf{P}(D)$ and $s := \frac{d}{r} - \lambda \in \mathbf{N}$ (i.e. the highest quotient of the composition series of $U^{(\lambda)}$-invariant spaces is unitarizable) is of particular interest.

THEOREM 3.4 *Let $\lambda \in \mathbf{P}(D)$ and assume that $s := \frac{d}{r} - \lambda \in \mathbf{N}$. Then, for each $\varphi \in Aut(D)$ and $f \in \mathcal{H}(D)$*

$$\partial_N^s(U^{(\lambda)}(\varphi) f) = U^{(p-\lambda)}(\varphi)(\partial_N^s f). \tag{3.8}$$

Namely, the operator ∂_N^s intertwines the actions $U^{(\lambda)}$ and $U^{(p-\lambda)}$ of $Aut(D)$. Moreover,

$$\langle f, g \rangle_{\lambda,q} = c_1 \langle \partial_N^s f, \partial_N^s g \rangle_{p-\lambda}, \quad \forall f, g \in \mathcal{H}_{\lambda,q}, \tag{3.9}$$

where

$$c_1^{-1} := (\frac{d}{r})_{(s,s,\ldots,s)} \prod_{j=1}^{r} {\prod_{u=0}^{s-1}}' (\lambda + u - \lambda_j), \tag{3.10}$$

and the product ${\prod'}_{u=0}^{s-1}$ *ranges over all non-zero terms. In particular, if* $\lambda < 1$*, then*

$$\langle f, g\rangle_{\lambda,q} = c_1\, c(p-\lambda) \int_D (\partial_N^s f)(z)\, \overline{(\partial_N^s g)(z)}\, h(z,z)^{-\lambda}\, dm(z), \quad \forall f, g \in \mathcal{H}_{\lambda,q}. \tag{3.11}$$

PROOF: (3.8) is proved in [A1], Theorem 6.4. For the proof of (3.9) and (3.11) we define an inner product on the polynomials modulo $\mathcal{M}_{q-1}^{(\lambda)}$ by

$$[f, g] := \langle \partial_N^s f, \partial_N^s g\rangle_{p-\lambda}, \quad f, g \in \mathcal{H}_{\lambda,q}.$$

We claim that $[\cdot,\cdot]$ is $U^{(\lambda)}$-invariant. Indeed, using (3.8) we see that for every $\varphi \in Aut(D)$ and polynomials f and g,

$$\begin{aligned} [U^{(\lambda)}(\varphi)f, U^{(\lambda)}(\varphi)g] &= \langle \partial_N^s(U^{(\lambda)}(\varphi)f), \partial_N^s(U^{(\lambda)}(\varphi)g)\rangle_{p-\lambda} \\ &= \langle U^{(p-\lambda)}(\varphi)(\partial_N^s f), U^{(p-\lambda)}(\varphi)(\partial_N^s g)\rangle_{p-\lambda} \\ &= \langle \partial_N^s f, \partial_N^s g\rangle_{p-\lambda} = [f, g]. \end{aligned}$$

Since polynomials are dense in $\mathcal{H}_{\lambda,q}$, the fact that its inner product is the unique $U^{(\lambda)}$-invariant inner product (see [AF], [A1]) implies that

$$\langle f, g\rangle_{\lambda,q} = c_1\, [f, g], \quad \forall f, g \in \mathcal{H}_{\lambda,q}.$$

The value (3.10) of c_1 is found by taking $f = g = N^s$, and using the facts that $\langle N^s, N^s\rangle_{\mathcal{F}} = (\frac{d}{r})_{(s,s,\ldots,s)}$, $[N^s, N^s] = (\partial_N^s N^s)^2 = \langle N^s, N^s\rangle_{\mathcal{F}}^2$, and

$$\langle N^s, N^s\rangle_{\lambda,q} = \lim_{\xi\to\lambda} (\xi - \lambda)^q \frac{\langle N^s, N^s\rangle_{\mathcal{F}}}{(\xi)_{(s,s\ldots,s)}} = \frac{\langle N^s, N^s\rangle_{\mathcal{F}}}{\prod_{j=1}^{r} {\prod'}_{u=0}^{s-1}(\lambda + u - \lambda_j)}. \qquad \blacksquare$$

EXAMPLE: In the special case where $\lambda = 0$ and $s := \frac{d}{r} \in \mathbf{N}$, $\mathcal{H}_{0,q}$ is the generalized Dirichlet space, and formula (3.11) is the generalized Dirichlet inner product

$$\langle f, g\rangle_{0,q} = c_1\, c(p-\lambda) \int_D (\partial_N^s f)(z)\, \overline{(\partial_N^s g)(z)}\, dm(z), \quad \forall f, g \in \mathcal{H}_{0,q}.$$

4 THE EXPANSION OF THE OPERATORS $D_\ell(\lambda)$

Yan's operators $D_\ell(\lambda) = N^{\frac{d}{r}-\lambda}\partial_N^\ell N^{\lambda+\ell-\frac{d}{r}}$ and their derivatives play an important role in the previous section. In this section we obtain an expansion of $D_\ell(\lambda)$ in powers of λ. This expansion will exhibit $D_\ell(\lambda)$ as a polynomial in z, $\frac{\partial}{\partial z}$, and λ, showing that $D_\ell(\lambda)$ is a differential operator (with parameters λ and ℓ) in the ordinary sense. It also facilitates the computation of the derivatives

$$D_\ell^\nu(\lambda) = \frac{1}{\nu!}(\frac{\partial}{\partial\xi})^\nu D_\ell(\xi)_{|\xi=\lambda},$$

needed in formulas (3.3), (3.5) and (3.6) for the forms $\langle f,g\rangle_{\lambda,j}$. Another consequence will be that for any r distinct complex numbers $\alpha_1,\ldots,\alpha_r$ the operators $D_1(\alpha_1),\ldots,D_1(\alpha_r)$ are algebraically independent generators of the ring of invariant differential operators on the cone Ω, a result obtained independently also by Korányi and Yan (see [FK2], Chapter XIV). We shall work in the framework of the cone Ω, but all the results will be valid for Z, because Ω is a set of uniqueness for analytic functions on Z.

EXAMPLE 4.1. Let $D\subset \mathbf{C}^d$, $d\geq 3$ be a Cartan domain of rank $r=2$ (called the *Lie ball*). The associated JB*-algebra $Z=\mathbf{C}^d$, called the *complex spin factor*, is defined via

$$zw := (z_1w_1 - z'\cdot w', z_1w' + w_1z'), \quad z^* := (\overline{z_1}, -\overline{z'}),$$

where $z=(z_1,z')$, $z'=(z_2,z_3,\ldots,z_d)$, and $z\cdot w := \sum_{j=1}^d z_jw_j$. The unit of Z is $e:=(1,0,0,\ldots,0)$, and the canonical frame is $\{e_1,e_2\}$, where $e_1 := \frac{1}{2}(1,i,0,0,\ldots,0)$, $e_2 := \frac{1}{2}(1,-i,0,0,\ldots,0)$. The norm polynomial and the associated differential operator are given by

$$N(z) := z\cdot z = \sum_{j=1}^d z_j^2 \quad \text{and} \quad \partial_N = N(\frac{\partial}{\partial z}) = \frac{1}{4}\sum_{j=1}^d \frac{\partial^2}{\partial z_j^2}$$

respectively, since $(z|w)=2z\cdot\overline{w}$ is the normalized inner product. Since $r=2$ and $a=d-2$, the Wallach set is

$$\mathbf{W}(D)=\mathbf{W}_d(D)\cup\mathbf{W}_c(D),\ \ \mathbf{W}_d(D)=\{0,\frac{d-2}{2}\},\ \ \mathbf{W}_c(D)=(\frac{d-2}{2},\infty).$$

One can show that D is given by

$$D=\{z\in Z;\left((\sum_{j=1}^d |z_j|^2)^2-|N(z)|^2\right)^{\frac{1}{2}} < 1-\sum_{j=1}^n |z_j|^2\}. \tag{4.1}$$

For every $\alpha\in\mathbf{C}$

$$\begin{aligned}\frac{\partial^2}{\partial z_k^2}N^\alpha &= \frac{\partial}{\partial z_k}(2\alpha N^{\alpha-1}z_k + N^\alpha\frac{\partial}{\partial z_k})\\ &= 2\alpha N^{\alpha-1}+4\alpha N^{\alpha-1}z_k\frac{\partial}{\partial z_k}+4\alpha(\alpha-1)N^{\alpha-2}z_k^2+N^\alpha\frac{\partial^2}{\partial z_k^2}.\end{aligned}$$

Since $R=\sum_{j=1}^d z_j\frac{\partial}{\partial z_j}$, we obtain

$$\partial_N N^\alpha = \frac{1}{4}(\sum_{j=1}^d\frac{\partial^2}{\partial z_j^2})N^\alpha = \alpha(\alpha-\frac{a}{2})N^{\alpha-1}+\alpha N^{\alpha-1}R+N^\alpha\partial_N.$$

It follows that for every $\alpha\in\mathbf{C}$ and $\ell\in\mathbf{N}$,

$$N^{1-\alpha}\partial_N N^\alpha = N\partial_N+\alpha R+\alpha(\alpha+\frac{d-2}{2})I = N\partial_N+(\alpha)_{(1,0)}R+(\alpha)_{(1,1)}I. \tag{4.2}$$

Since

$$D_\ell(\lambda) = \left(N^{\frac{d}{r}-\lambda}\partial_N N^{1+\lambda-\frac{d}{r}}\right)\left(N^{\frac{d}{r}-\lambda-1}\partial_N N^{2+\lambda-\frac{d}{r}}\right)\cdots\left(N^{\frac{d}{r}+1-\ell-\lambda}\partial_N N^{\ell+\lambda-\frac{d}{r}}\right),$$

we finally obtain

$$D_\ell(\lambda) = \prod_{j=1}^{\ell}(N\partial_N + (\lambda - \frac{d}{2} + j)R + (\lambda - 1 + j)(\lambda - \frac{d}{2} + j)I). \tag{4.3}$$

Note that the factors on the right hand sides of (4.2) and (4.3) commute, since they are $G(\Omega)$-invariant, and the entire ring of $G(\Omega)$-invariant operators is commutative. Also, the operators R and $N\partial_N$ are K-invariant. Hence the factors on the right hand sides of (4.2) and (4.3) are multipliers of the Peter-Weyl decomposition of analytic functions on D (see Corollary 2.1).

Consider a general Cartan domain of tube-type $D \subset \mathbf{C}^d$ with rank r. Let Ω be the associated symmetric cone in the Euclidean Jordan algebra X and fix a frame $\{e_1, \ldots, e_r\}$ of pairwise orthogonal primitive idempotents in X, whose sum is the unit element e. For $1 \leq \nu \leq r$, let $\phi_\nu := \phi_{\mathbf{1}_\nu}$ be the spherical polynomial associated with the signature $\mathbf{1}_\nu := (1, 1, \ldots, 1, 0, 0, \ldots, 0)$, where there are ν "1"'s and $r - \nu$ "0"'s. Put also $\phi_0(z) \equiv 1$. Let $\{\Delta_\nu\}_{\nu=0}^r$ be the differential operators on Ω defined via

$$(\Delta_\nu)f(a) := \phi_\nu(\frac{d}{dx})(f(P(a^{\frac{1}{2}})x))_{|x=e}, \tag{4.4}$$

where for $b \in X$, $P(b)$ is defined via (1.1). Recall that $P(b) \in G(\Omega)$ for every $b \in \Omega$, and that $\Omega = \{P(b)e; b \in \Omega\}$ since $P(a^{\frac{1}{2}})e = a$. Moreover, the L-invariance of the ϕ_ν's and the "polar decomposition" for Ω imply that

$$(\Delta_\nu)f(a) := \phi_\nu(\frac{d}{dx})(f(\psi(x)))_{|x=e}, \quad a \in \Omega \tag{4.5}$$

for every $\psi \in G(\Omega)$ for which $\psi(e) = a$. This implies that the operators $\{\Delta_\nu\}_{\nu=0}^r$ are $G(\Omega)$-invariant, namely

$$\Delta_\nu(f \circ \psi) = (\Delta_\nu f) \circ \psi, \quad \forall \psi \in G(\Omega), \quad \forall f \in C^\infty(\Omega).$$

We remark that (4.4) and (4.5) are equivalent to

$$\Delta_\nu e^{\langle x,y\rangle}{}_{|x=a} = \phi_\nu(\psi^*(y))\, e^{\langle a,y\rangle} = \phi_\nu(P(a^{\frac{1}{2}})y)\, e^{\langle a,y\rangle}, \ \ a, y \in \Omega, \tag{4.6}$$

where $\psi \in G(\Omega) \subset GL(X)$ satisfies $\psi(e) = a$, ψ^* is the adjoint of ψ with respect to the inner product $\langle\cdot,\cdot\rangle$ on X, and Δ_ν differentiates the coordinate x. Notice also that the operators Δ_ν can be written as

$$\Delta_\nu = c_{\mathbf{m}} K_{\mathbf{m}}(x, \frac{\partial}{\partial x}),$$

where $\mathbf{m} = (1, 1, \ldots, 1, 0, \ldots, 0)$ (ν "ones" and $r - \nu$ zeros), and $c_{\mathbf{m}}$ is an appropriate constant.

For $\nu = 0, 1, r$ it is easy to compute Δ_ν. Clearly, $\Delta_0 = I$. Since N is L-invariant, $\phi_r = N$. Using (4.6) and (1.3), we find that

$$\Delta_r = N\,\partial_N.$$

Also, $\phi_1(x) = \frac{1}{r}\,tr(x) = \frac{1}{r}\,\langle x, e\rangle$. Indeed, using $N_1(x) = \langle x, e_1\rangle$ and the fact that L is transitive on the frames, we get

$$\begin{aligned} \phi_1(x) &= \int_L \langle \ell x, e_1\rangle\, d\ell = \frac{1}{r}\sum_{j=1}^r \int_L \langle \ell x, e_j\rangle\, d\ell \\ &= \frac{1}{r}\int_L \langle \ell x, e\rangle\, d\ell = \frac{1}{r}\int_L \langle x, \ell e\rangle\, d\ell = \frac{1}{r}\langle x, e\rangle. \end{aligned}$$

Using the fact that $tr(P(a^{\frac{1}{2}})y) = \langle P(a^{\frac{1}{2}})y, e\rangle = \langle y, P(a^{\frac{1}{2}})e\rangle = \langle y, a\rangle, \quad \forall a, y \in \Omega$, we find that

$$\Delta_1 = \frac{1}{r}R,$$

where $Rf(x) := \frac{\partial}{\partial t} f(tx)_{|t=1}$ is the radial derivative.

Our main result in this section is the expansion of $D_1(\lambda) = N^{\frac{d}{r}-\lambda}\,\partial_N\, N^{1+\lambda-\frac{d}{r}}$. This result was obtained independently by A. Korányi, see [FK2], Proposition XIV.1.5.

Theorem 4.1 *For every* $\lambda \in \mathbf{C}$,

$$D_1(\lambda) = \sum_{\nu=0}^r \binom{r}{\nu} \prod_{j=\nu+1}^r (\lambda - \lambda_j)\,\Delta_\nu. \tag{4.7}$$

Proof: For $x \in \Omega$, the function $\alpha \to N(x)^\alpha$ is entire in α. Hence both sides of (4.7) are entire in λ, and it is therefore enough to prove (4.7) for λ with $\Re\lambda < 0$. Let $\alpha = \lambda_r - \lambda$. Since $\Re\lambda > \lambda_r$, we get for every $x \in \Omega$

$$N(x)^{-\alpha} = \frac{1}{\Gamma_\Omega(\alpha)} \int_\Omega e^{-\langle x,t\rangle} N(t)^\alpha\, d\mu_\Omega(t),$$

where $d\mu_\Omega(t) := N(t)^{-\frac{d}{r}}dt$ is the $G(\Omega)$-invariant measure on Ω. Fix $a, y \in \Omega$ and put $f_y(x) := e^{\langle x,y\rangle}$. Then

$$\begin{aligned} &(N^{\alpha+1}\partial_N N^{-\alpha} f_y)(a) \\ &= \frac{N(a)^{\alpha+1}}{\Gamma_\Omega(\alpha)} N(\frac{d}{dx}) \int_\Omega e^{\langle x,y-t\rangle} N(t)^\alpha d\mu_\Omega(t)\,_{|x=a} \\ &= \frac{N(a)^{\alpha+1}}{\Gamma_\Omega(\alpha)} \int_\Omega e^{\langle a,y-t\rangle} N(y-t) N(t)^\alpha\, d\mu_\Omega(t) \\ &= \frac{f_y(a)}{\Gamma_\Omega(\alpha)} \int_\Omega e^{-\langle e, P(a^{\frac{1}{2}})t\rangle} N(P(a^{\frac{1}{2}})(y-t)) N(P(a^{\frac{1}{2}})t)^\alpha\, d\mu_\Omega(t). \end{aligned}$$

Letting $b = P(a^{\frac{1}{2}})y$, the substitution $t := P(a^{-\frac{1}{2}})P(b^{\frac{1}{2}})\tau$ gives

$$(N^{\alpha+1}\partial_N N^{-\alpha} f_y)(a) = \frac{f_y(a)}{\Gamma_\Omega(\alpha)} N(y)^{1+\alpha} N(a)^{1+\alpha} \int_\Omega e^{-\langle b,\tau\rangle} N(e-\tau) N(\tau)^\alpha \, d\mu_\Omega(\tau).$$

Now, the well-known "binomial formula"

$$N(e+x) = \sum_{\nu=0}^{r} \binom{r}{\nu} \phi_\nu(x), \quad x \in X \tag{4.8}$$

(which follows from Theorem 1.2 and the knowledge of the norms of the ϕ_ν's) and the fact that for every $\mathbf{s} \in \mathbf{C}^r$ and $b \in \Omega$

$$\frac{1}{\Gamma_\Omega(\mathbf{s})} \int_\Omega e^{-\langle b,\tau\rangle} \phi_{\mathbf{s}}(\tau) \, d\mu_\Omega(\tau) = \phi_{\mathbf{s}}(b^{-1}) \tag{4.9}$$

(which follows from the analogous formula for the conical functions), imply

$$\int_\Omega e^{-\langle b,\tau\rangle} N(e-\tau) N(\tau)^\alpha \, d\mu(\tau) = \sum_{\nu=0}^{r} \binom{r}{\nu} \int_\Omega e^{-\langle b,\tau\rangle} \phi_{\mathbf{1}_\nu+\alpha}(\tau) \, d\mu_\Omega(\tau)$$

$$= \sum_{\nu=0}^{r} \binom{r}{\nu} \Gamma_\Omega(\mathbf{1}_\nu+\alpha) \, \phi_{\mathbf{1}_\nu+\alpha}(b^{-1}) = N(b)^{-\alpha} \sum_{\nu=0}^{r} \binom{r}{\nu} \Gamma_\Omega(\mathbf{1}_\nu+\alpha) \, \phi_\nu(b^{-1}).$$

We claim that for every $b \in \Omega$ and $1 \le \nu \le r$,

$$\phi_\nu(b^{-1}) = \phi_{r-\nu}(b) \, N(b)^{-1}. \tag{4.10}$$

Indeed, using (4.8) we have $N(e+tb^{-1}) = \sum_{\nu=0}^{r} \binom{r}{\nu} \phi_\nu(b^{-1}) \, t^\nu$, as well as

$$\begin{aligned} N(e+tb^{-1}) &= N(P(b^{-\frac{1}{2}})(b+te)) = N(b)^{-1} t^r \, N(e+t^{-1}b) \\ &= N(b)^{-1} t^r \sum_{k=0}^{r} \binom{r}{k} \phi_k(b) \, t^{-k}. \end{aligned}$$

Comparing the coefficients of t^ν in the two expansions, we obtain (4.10). It follows that

$$\begin{aligned} &(N^{\alpha+1} \partial_N N^{-\alpha} f_y)(a) \\ &= \frac{f_y(a) \, N(y)^{1+\alpha} \, N(a)^{1+\alpha}}{\Gamma_\Omega(\alpha) \, N(b)^{1+\alpha}} \sum_{\nu=0}^{r} (-1)^\nu \binom{r}{\nu} \Gamma_\Omega(\mathbf{1}_\nu+\alpha) \, \phi_{r-\nu}(b) \\ &= f_y(a) \sum_{\nu=0}^{r} (-1)^\nu \binom{r}{\nu} \frac{\Gamma_\Omega(\mathbf{1}_\nu+\alpha)}{\Gamma_\Omega(\alpha)} \phi_{r-\nu}(b) \\ &= f_y(a) \sum_{\nu=0}^{r} \binom{r}{\nu} \prod_{j=1}^{\nu} (\lambda_j - \alpha) \, \phi_{r-\nu}(P(a^{\frac{1}{2}})y). \end{aligned}$$

Comparing this with (4.6), we conclude that

$$N^{\alpha+1}\,\partial_N N^{-\alpha} = \sum_{\nu=0}^{r} \binom{r}{\nu} \prod_{j=1}^{\nu}(\lambda_j-\alpha)\;\Delta_{r-\nu} = \sum_{k=0}^{r}\binom{r}{k}\prod_{j=1}^{r-k}(\lambda_j-\alpha)\;\Delta_k .$$

Using the relations $\alpha = \lambda_r - \lambda$ and $\frac{d}{r} = 1+\lambda_r$, we obtain (4.7). ∎

REMARK: The "binomial formula" (4.8) yields that for every $\nu = 1, 2, \ldots, r$ and every $x \in X$,

$$\phi_\nu(x) = \sum_{1\le i_1<i_2<\ldots<i_\nu\le r} \lambda_{i_1}\,\lambda_{i_2}\,\cdots\,\lambda_{i_\nu}/\binom{r}{\nu} = S_{r,\nu}(\lambda)/\binom{r}{\nu},$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ is the sequence of eigenvalues of x, and $S_{r,\nu}$ is the elementary symmetric polynomial of degree ν in r variables.

Combining the definition $D_\ell(\lambda) = \prod_{k=0}^{\ell-1} D_1(\lambda+k)$ with Theorem 4.1, we obtain

COROLLARY 4.1 *For every $\lambda \in \mathbf{C}$ and $\ell \in \mathbf{N}$,*

$$D_\ell(\lambda) = \prod_{k=0}^{\ell-1}\sum_{\nu=0}^{r}\binom{r}{\nu}\prod_{j=\nu+1}^{r}(\lambda+k-\lambda_j)\;\Delta_\nu . \tag{4.11}$$

For any signature $\mathbf{m} \ge 0$ let $\Delta_{\mathbf{m}}$ be the differential operator associated with the spherical polynomial $\phi_{\mathbf{m}}$ via

$$(\Delta_{\mathbf{m}} f)(a) := \phi_{\mathbf{m}}(\frac{d}{dx})\, f(P(a^{\frac{1}{2}}))_{|x=e}, \quad a\in\Omega. \tag{4.12}$$

Equivalently,

$$\Delta_{\mathbf{m}} e^{\langle x,y\rangle}{}_{|x=a} = \phi_{\mathbf{m}}(P(a^{\frac{1}{2}})y)\; e^{\langle a,y\rangle}, \quad a\in\Omega. \tag{4.13}$$

Again, one can replace in (4.12) and (4.13) $P(a^{\frac{1}{2}})$ by any $\psi \in G(\Omega)$ satisfying $\psi(e) = a$. Hence the operators $\Delta_{\mathbf{m}}$ are $G(\Omega)$-invariant, namely

$$\Delta_{\mathbf{m}}(f\circ\psi) = (\Delta_{\mathbf{m}} f)\circ\psi, \quad \forall\psi\in G(\Omega).$$

THEOREM 4.2 *For every $\lambda \in \mathbf{C}$ and $\ell \in \mathbf{N}$,*

$$\begin{aligned} D_\ell(\lambda) &= \sum_{\mathbf{m}\ge 0}^{(\ell)} \frac{\Gamma_\Omega(\frac{d}{r}+\ell)\,\Gamma_\Omega(\frac{d}{r}-\lambda-\mathbf{m}^*)}{\Gamma_\Omega(\frac{d}{r}+\ell-\mathbf{m}^*)\,\Gamma_\Omega(\frac{d}{r}-\ell-\lambda)}\;\frac{d_{\mathbf{m}}}{(\frac{d}{r})_{\mathbf{m}}}\;\Delta_{\mathbf{m}} \\ &= (\frac{d}{r}-\lambda-\ell)_{(\ell,\ldots,\ell)}\sum_{\mathbf{m}\ge 0}^{(\ell)}\frac{(-\ell)_{\mathbf{m}}}{(\lambda)_{\mathbf{m}}}\;\frac{d_{\mathbf{m}}}{(\frac{d}{r})_{\mathbf{m}}}\;\Delta_{\mathbf{m}}. \end{aligned} \tag{4.14}$$

Here $\mathbf{m}^* := (m_r, m_{r-1}, \ldots, m_1)$, $d_{\mathbf{m}} = dim(P_{\mathbf{m}})$, and the summation $\sum_{\mathbf{m}\ge 0}^{(\ell)}$ extends over all $\mathbf{m} = (m_1, m_2, \ldots, m_r) \in \mathbf{N}^r$ with $\ell \ge m_1 \ge m_2 \ge \ldots \ge m_r \ge 0$.

Proof: The general binomial formula (1.15) and the relations

$$K_{\mathbf{m}}(x,e)=\frac{\phi_{\mathbf{m}}}{\|\phi_{\mathbf{m}}\|_F^2},\qquad \|\phi_{\mathbf{m}}\|_F^2=\frac{(\frac{d}{r})_{\mathbf{m}}}{d_{\mathbf{m}}}$$

(see [FK2], Chapter XI) imply for $\ell\in\mathbf{N}$ and $x\in X$

$$N(e+x)^\ell = c\sum_{\mathbf{m}\geq 0}^{(\ell)}\frac{d_{\mathbf{m}}}{(\frac{d}{r})_{\ell-\mathbf{m}^*}\,(\frac{d}{r})_{\mathbf{m}}}\,\phi_{\mathbf{m}}(x), \tag{4.15}$$

where $c:=(\frac{d}{r})_{(\ell,\ell,\ldots,\ell)}$, and $\mathbf{m}^*$ and $\sum_{\mathbf{m}\geq 0}{}^{(\ell)}$ are as in Theorem 4.2. Indeed, by (1.15),

$$N(e+x)^\ell=\sum_{\mathbf{m}\geq 0}(-\ell)_{\mathbf{m}}\frac{(-1)^{|\mathbf{m}|}\,d_{\mathbf{m}}}{(\frac{d}{r})_{\mathbf{m}}}\,\phi_{\mathbf{m}}(x).$$

From this (4.15) follows by the fact that $(-\ell)_{\mathbf{m}}=0$ if $m_1>\ell$, whereas in case $m_1\leq\ell$,

$$(-\ell)_{\mathbf{m}}\,(-1)^{|\mathbf{m}|}=\frac{(\frac{d}{r})_{(\ell,\ell,\ldots,\ell)}}{(\frac{d}{r})_{\ell-\mathbf{m}^*}}.$$

As in the proof of Theorem 4.1, it is enough to prove that for every $\alpha\in\mathbf{C}$ with $\Re\alpha>\lambda_r$ and every $\ell\in\mathbf{N}$,

$$N^{\alpha+\ell}\,\partial_N^\ell\,N^{-\alpha}=c\sum_{\mathbf{m}\geq 0}^{(\ell)}\frac{(\alpha)_{\ell-\mathbf{m}^*}\,d_{\mathbf{m}}}{(\frac{d}{r})_{\ell-\mathbf{m}^*}\,(\frac{d}{r})_{\mathbf{m}}}\,\Delta_{\mathbf{m}}. \tag{4.16}$$

From this one obtains (4.14) by the substitution $\alpha=\frac{d}{r}-\ell-\lambda$. To prove (4.16), fix $a,y\in\Omega$ and let $f_y(x):=e^{\langle x,y\rangle}$. Then

$$\begin{aligned}(N^{\alpha+\ell}\,\partial_N^\ell\,N^{-\alpha}f_y)(a) &= \frac{N(a)^{\alpha+\ell}\,f_y(a)}{\Gamma_\Omega(\alpha)}\int_\Omega e^{-\langle a,t\rangle}\,N(y-t)^\ell\,N(t)^\alpha\,d\mu_\Omega(t)\\ &= \frac{N(b)^{\alpha+\ell}\,f_y(a)}{\Gamma_\Omega(\alpha)}\int_\Omega e^{-\langle b,u\rangle}\,N(e-u)^\ell\,N(u)^\alpha\,d\mu_\Omega(u),\end{aligned}$$

by the substitutions $b=P(a^{\frac{1}{2}})y$ and $u=P(b^{-\frac{1}{2}})P(a^{\frac{1}{2}})t$. Using (4.15), (4.9), and

$$\phi_{\mathbf{m}}(x^{-1})=\phi_{\ell-\mathbf{m}^*}(x)\,N(x)^{-\ell} \tag{4.17}$$

(a consequence of [FK2], Proposition VII.1.5), we obtain

$$(N^{\alpha+\ell}\,\partial_N^\ell\,N^{-\alpha}f_y)(a)=c\,\frac{f_y(a)}{\Gamma_\Omega(\alpha)}\sum_{\mathbf{m}\geq 0}^{(\ell)}\frac{\Gamma_\Omega(\mathbf{m}+\alpha)\,d_{\mathbf{m}}}{(\frac{d}{r})_{\ell-\mathbf{m}^*}\,(\frac{d}{r})_{\mathbf{m}}}\,\phi_{\ell-\mathbf{m}^*}(P(a^{\frac{1}{2}})y).$$

With the change of variables $\mathbf{n}:=\ell-\mathbf{m}^*$, the fact that $d_{\mathbf{m}}=d_{\mathbf{n}}$ (use (4.17) or the general formula for $d_{\mathbf{m}}$ in [U1]), the definition (4.12), and

$$(N^{\alpha+\ell}\,\partial_N^\ell\,N^{-\alpha}f_y)(a)=c\,f_y(a)\sum_{\mathbf{n}\geq 0}^{(\ell)}\frac{(\alpha)_{\ell-\mathbf{n}^*}\,d_{\mathbf{n}}}{(\frac{d}{r})_{\ell-\mathbf{n}^*}\,(\frac{d}{r})_{\mathbf{n}}}\,\phi_{\mathbf{n}^*}(P(a^{\frac{1}{2}})y),$$

we obtain (4.16). ∎

COROLLARY 4.2 *The operators $\{\Delta_k\}_{k=1}^r$ are algebraically independent generators of the ring Diff$(\Omega)^{G(\Omega)}$ of $G(\Omega)$-invariant differential operators on Ω.*

PROOF: Comparing the two expansions (4.11) and (4.14) of $D_\ell(\lambda)$, we see that

$$\Delta_{\mathbf{m}} \in \mathbf{C}[\Delta_1, \Delta_2, \ldots, \Delta_r]$$

for every signature $\mathbf{m} \geq 0$. Since $\{\phi_{\mathbf{m}}\}_{\mathbf{m}\geq 0}$ is a basis for the space of spherical polynomials, the one-to-one correspondence between spherical polynomials and the elements of Diff$(\Omega)^{G(\Omega)}$ (see [FK2], Chapter XIV) implies that $\{\Delta_{\mathbf{m}}\}_{\mathbf{m}\geq 0}$ is a basis of Diff$(\Omega)^{G(\Omega)}$. Thus Diff$(\Omega)^{G(\Omega)} = \mathbf{C}[\Delta_1, \Delta_2, \ldots, \Delta_r]$. Since the minimal number of algebraic generators of Diff$(\Omega)^{G(\Omega)}$ is $r = rank(\Omega)$ [He], it follows that $\Delta_1, \Delta_2, \ldots, \Delta_r$ are algebraically independent. ∎

The *divided differences* of a C^1-function f on $\mathbf{R}$ are defined by

$$f^{[1]}(t_0, t_1) := \frac{f(t_0) - f(t_1)}{t_0 - t_1}$$

for $t_0 \neq t_1$, and $f^{[1]}(t_0, t_0) := f'(t_0)$. The higher order divided differences of a smooth enough function f are defined inductively by

$$f^{[n]}(t_0, t_1, \ldots, t_n) := g^{[1]}(t_{n-1}, t_n),$$

where $g(x) := f^{[n-1]}(t_0, t_1, \ldots, t_{n-2}, x)$. Then $f^{[n]}(t_0, t_1, \ldots, t_n)$ is symmetric in $t_0, t_1, \ldots, t_n$, and

$$f^{[n]}(t, t, \ldots, t) = \frac{1}{n!} \frac{d^n}{dt^n} f(t).$$

Moreover, if f is analytic in a domain $\mathcal{D} \subset \mathbf{C}$, then

$$f^{[n]}(t_0, t_1, \ldots, t_n) = \frac{1}{2\pi i} \int_\Gamma \frac{f(\xi)}{\prod_{j=0}^n (\xi - t_j)}\, d\xi$$

for all $t_0, t_1, \ldots, t_n \in \mathcal{D}$ and every Jordan curve Γ in $\mathcal{D}$ whose interior contains $t_0, t_1, \ldots, t_n$ and is contained in $\mathcal{D}$. The divided differences of vector-valued maps are defined in the same way and have analogous properties. For convenience we put also $f^{[0]}(t) := f(t)$.

THEOREM 4.3 *Let $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbf{C}$ be distinct. Then $\{D_1(\alpha_j)\}_{j=1}^r$ are algebraically independent generators of Diff$(\Omega)^{G(\Omega)}$. Moreover, for $\ell = 1, 2, \ldots, r$,*

$$\Delta_\ell = D_1^{[r-\ell]}(\lambda_\ell, \lambda_{\ell+1}, \ldots, \lambda_r) / \binom{r}{\nu}, \tag{4.18}$$

where $D_1^{[r-\ell]}(\lambda_\ell, \ldots, \lambda_r)$ are the divided differences of order $r - \ell$ of $D_1(\lambda)$, evaluated at $(\lambda_\ell, \lambda_{\ell+1}, \ldots, \lambda_r)$.

Proof: Let $h_k(x) := \binom{r}{\ell} \prod_{j=k+1}^{r}(x-\lambda_j)$, $0 \le k \le r$. Then $h_k^{[m]}(x_0, x_1, \ldots, x_m) \equiv 0$ whenever $m > r-k$, and $h_k^{[r-k]}(x_0, x_1, \ldots, x_{r-k}) \equiv \binom{r}{\ell}$ for all choices of $x_0, x_1, \ldots, x_{r-k}$. By Theorem 4.2, $D_1(\alpha) = \sum_{k=0}^{r} h_k(\alpha)\, \Delta_k$. Hence, for $1 \le \ell \le r$,

$$D_1^{[r-\ell]}(\alpha_\ell, \alpha_{\ell+1}, \ldots, \alpha_r) = \sum_{k=0}^{\ell} h_k^{[r-\ell]}(\alpha_\ell, \alpha_{\ell+1}, \ldots, \alpha_r)\, \Delta_k.$$

Solving this system of equations for the Δ_k's, we see that $\mathrm{Diff}(\Omega)^{G(\Omega)} = \mathbf{C}[\Delta_1, \Delta_2, \ldots, \Delta_r]$ coincides with the ring generated by the operators $\{D_1^{[r-\ell]}(\alpha_\ell, \alpha_{\ell+1}, \ldots, \alpha_r)\}_{\ell=1}^{r}$. If the $\{\alpha_j\}_{j=1}^{r}$ are distinct, then

$$D_1^{[r-\ell]}(\alpha_\ell, \alpha_{\ell+1}, \ldots, \alpha_r) \;\in\; \mathbf{C}[D_1(\alpha_1), D_1(\alpha_2), \ldots, D_1(\alpha_r)].$$

Hence,

$$\mathrm{Diff}(\Omega)^{G(\Omega)} \;=\; \mathbf{C}[\Delta_1, \Delta_2, \ldots, \Delta_r] \;=\; \mathbf{C}[D_1(\alpha_1), D_1(\alpha_2), \ldots, D_1(\alpha_r)].$$

The operators $\{D_1(\alpha_j\}_{j=1}^{r}$ are algebraically independent, since $\mathrm{Diff}(\Omega)^{G(\Omega)}$ cannot be algebraically generated by less than r elements. If $\alpha_j = \lambda_j$ for $j = 1, 2, \ldots, r$, then $h_k^{[r-\ell]}(\alpha_\ell, \ldots, \alpha_r) = 0$ for $k < \ell$. Thus, for $\ell = 1, 2, \ldots, r$,

$$D_1^{[r-\ell]}(\alpha_\ell, \alpha_{\ell+1}, \ldots, \alpha_r) \;=\; h_\ell^{[r-\ell]}(\alpha_\ell, \alpha_{\ell+1}, \ldots, \alpha_r)\, \Delta_\ell = \binom{r}{\ell} \Delta_\ell. \qquad \blacksquare$$

Remark: The first statement in Theorem 4.3 was proved independently also by A. Korányi [FK2] and Z. Yan [Y1]. Our result is slightly stronger, giving the exact formula (4.18).

Combining Theorems 3.2 and 4.2 (or, 4.1) we obtain integral formulas for the invariant hermitian forms $\langle\cdot,\cdot\rangle_{\lambda,j}$, $\lambda \in \mathbf{P}(D)$, $0 \le j \le q(\lambda)$.

Corollary 4.3 *Let $\lambda \in \mathbf{P}(D)$, $\ell \in \mathbf{N}$ and assume that $\lambda + \ell \ge \frac{d}{r} = \lambda_r + 1$. Let $q = q(\lambda)$, $0 \le j \le q$, and $\nu = q - j$. Consider the $G(\Omega)$-invariant differential operator*

$$T_{\lambda,j} := \gamma \sum_{\mathbf{m}\ge 0}^{(\ell)} c_{\mathbf{m}}(\lambda,\ell)\, \frac{d_{\mathbf{m}}}{(\frac{d}{r})_{\mathbf{m}}}\; \Delta_{\mathbf{m}}, \tag{4.19}$$

where γ is given by (3.4), and for every $\mathbf{m} \ge 0$ with $m_1 \le \ell$

$$c_{\mathbf{m}}(\lambda,\ell) := \frac{1}{\nu!}\Big(\frac{\partial}{\partial\xi}\Big)^{\ell} \left(\frac{\Gamma_\Omega(\frac{d}{r}+\ell)\, \Gamma_\Omega(\frac{d}{r}-\xi-\mathbf{m}^*)}{\Gamma_\Omega(\frac{d}{r}+\ell-\mathbf{m}^*)\, \Gamma_\Omega(\frac{d}{r}-\ell-\xi)} \right)_{|\xi=\lambda}. \tag{4.20}$$

Then $T_{\lambda,j}$ is defined on all analytic functions on D, and for all $f, g \in \mathcal{H}_{\lambda,j}$

$$\langle f, g\rangle_{\lambda,j} \;=\; \langle T_{\lambda,j} f, g\rangle_{\lambda+\ell}. \tag{4.21}$$

In particular, if $\lambda + \ell > p - 1$ or $\lambda + \ell = \frac{d}{r}$ then we have

$$\langle f, g\rangle_{\lambda,j} \;=\; \int_D (T_{\lambda,j} f)(z)\, \overline{g(z)}\; d\mu_{\lambda+\ell}(z) \;\; \textit{and} \;\; \langle f, g\rangle_{\lambda,j} \;=\; \int_S (T_{\lambda,j} f)(\xi)\, \overline{g(\xi)}\; d\sigma(\xi) \tag{4.22}$$

respectively.

The case $\lambda = \lambda_r$ is particularly simple, since then $\frac{d}{r} - \lambda_r = 1$, and we can use (4.7) rather than (4.14).

COROLLARY 4.4 *Let D be a Cartan domain of tube type and rank $r \geq 2$ in $\mathbf{C}^d$, $d \geq 3$. Then*

$$\langle f, g\rangle_{\lambda_r,0} = \langle \beta \sum_{\nu=0}^{r-1} \binom{r}{\nu} \prod_{i=2}^{r-\nu} \lambda_i \; \Delta_\nu f, g\rangle_{H^2(S)}, \quad where \quad \beta := \prod_{i=2}^{r} \lambda i. \tag{4.23}$$

PROOF: In this case $q = q(\lambda_r) = 1$, $j = 0$, and $\nu = q - j = 1$. We choose $\ell = 1$, so $\lambda_r + \ell = \frac{d}{r}$. In order to apply Theorem 3.2 we use Theorem 4.1, and compute

$$\begin{aligned} D_1^1(\lambda_r) = \frac{\partial}{\partial \xi} D_1(\xi)_{|\xi=\lambda} &= \frac{\partial}{\partial \xi} \left(\sum_{\nu=0}^{r} \binom{r}{\nu} \prod_{i=\nu+1}^{r} (\xi - \lambda_i)\, \Delta_\nu \right)_{|\xi=\lambda_r} \\ &= \sum_{\nu=0}^{r-1} \binom{r}{\nu} \prod_{i=\nu+1}^{r-1} (\lambda_r - \lambda_i)\, \Delta_\nu = \sum_{\nu=0}^{r-1} \binom{r}{\nu} \prod_{i=2}^{r-\nu} \lambda_i \; \Delta_\nu. \end{aligned}$$

Using this, (4.23) follows from

$$\beta := \frac{\partial}{\partial \xi} \left(\prod_{i=1}^{r} (\xi - \lambda_i) \right)_{\xi=\lambda} = \prod_{i=1}^{r-1} (\lambda_r - \lambda_i) = \prod_{i=2}^{r} \lambda_i. \qquad \blacksquare$$

EXAMPLE 4.2. Let D be the Cartan domain of rank $r = 2$ in $\mathbf{C}^d$ (the Lie ball), $d \geq 3$. Then

$$\langle f, g\rangle_{\frac{d-2}{2},0} = \langle (\frac{2}{d-2} R + I) f, g\rangle_{H^2(S)}. \tag{4.24}$$

Namely, in this case $\lambda = \lambda_2 = \frac{d-2}{2}$, $q = q(\lambda) = 1$, $j = 0$, and $\nu = q - j = 1$. With $\ell = 1$, $\lambda + \ell = \frac{d}{2} = \lambda_2 + 1 = \frac{d}{r}$ we get by using Theorem 3.2 and Corollary 3.2,

$$\begin{aligned} \langle f, g\rangle_{\frac{d-2}{2},0} &= \gamma \langle D_1^1(\frac{d-2}{2}) f, g\rangle_{\frac{d}{2}} \\ &= \gamma \, \langle (R + \frac{d-2}{2} I) f, g\rangle_{H^2(S)} = \langle (\frac{2}{d-2} R + I) f, g\rangle_{H^2(S)}. \end{aligned}$$

Since the Shilov boundary S of D is given by

$$S = \{e^{i\theta}(x_1, ix_2, ix_3, \ldots, ix_d); \theta \in \mathbf{R}, \sum_{j=1}^{d} x_j^2 = 1\} \equiv S^1 \cdot S^{d-1},$$

the unique K-invariant probability measure on S is $d\sigma(e^{i\theta}(x_1, ix')) = \frac{d\theta}{2\pi} d\nu_{d-1}(x)$, where ν_{d-1} is the unique $O(d-1)$-invariant probability measure on S^{d-1}. Thus (4.24) provides a very concrete formula for the inner product $\langle \cdot, \cdot \rangle_{\frac{d-2}{2},0}$.

5 INTEGRATION OVER BOUNDARY ORBITS OF $Aut(D)$

In this section we obtain formulas for the invariant inner products in terms of integration over an orbit of $Aut(D)$ on the boundary ∂D. We focus on the inner products $\langle\cdot,\cdot\rangle_{\lambda_2,0} = \langle\cdot,\cdot\rangle_{\frac{a}{2}}$, and conjecture that our method can be generalized for the derivation of similar formulas for the inner products $\langle\cdot,\cdot\rangle_{\lambda_j,0} = \langle\cdot,\cdot\rangle_{\lambda_j}$, $\lambda_j = (j-1)\frac{a}{2}$, $j = 3,4,\ldots,r$, in terms of integration on an appropriate boundary orbit. (Notice that the case $j=1$ is trivial, since $\lambda_1 = 0$ and $\mathcal{H}_{0,0} = \mathcal{H}_0 = \mathbf{C}1$).

In order to describe the *facial structure* of a Cartan domain of tube-type $D \subset \mathbf{C}^d$ [Lo], [A1], let S_ℓ be the compact, real analytic manifold of tripotents in Z of rank $\ell = 1,2,\ldots,r$. The group K acts transitively and irreducibly on S_ℓ. Let σ_ℓ be the unique K-invariant probability measure on S_ℓ given by

$$\int_{S_\ell} f\, d\sigma_\ell = \int_K f(k(v_\ell))\, dk, \tag{5.1}$$

where v_ℓ is any fixed element of S_ℓ. For any tripotent v let $Z = Z_1(v) + Z_{\frac{1}{2}}(v) + Z_0(v)$ be the corresponding Peirce decomposition. Then $D_v := D \cap Z_0(v)$ is a Cartan domain of tube-type, which is the open unit ball of the JB*-algebra $Z_0(v)$. If $v \in S_\ell$ then the rank of D_v is $r_v := r - \ell$, its characteristic multiplicity is $a_v := a$ if $\ell \le r-2$ and $a_v = 0$ if $\ell = r-1$, and the genus is $p_v = p - \ell a$. The set $v + D_v$ is a face of the closure $\overline{D}$ of D. For any function f on $\overline{D}$ let f_v be the function on $\overline{D_v}$ defined by

$$f_v(z) := f(v+z), \quad z \in \overline{D_v}. \tag{5.2}$$

The fundamental polynomial "h" of $Z_0(v)$ is defined by

$$h_v(z,w) := h(z,w), \quad z,w \in Z_0(v). \tag{5.3}$$

For $\ell = 1,2,\ldots,r$, $\partial_\ell D := \cup_{v\in S_\ell}(v + D_v)$ is an orbit of G: $\partial_\ell D = G(v_\ell)$. If $v \in S_r$ is a maximal tripotent, then $D_v = Z_0(v) = \{0\}$. Hence $\partial_r D = S_r = S$ is the Shilov boundary. In particular, S is a G-orbit. The only tripotent of rank 0 is $0 \in Z$, and $D = D_0$ is also a G-orbit. Thus the decomposition of $\overline{D}$ into G-orbits is

$$\overline{D} = D \cup \bigcup_{\ell=1}^{r} \partial_\ell D.$$

For every tripotent $v \in Z$ and $\lambda > p_v - 1$ consider the probability measure $\mu_{v,\lambda}$ on D_v, defined via

$$\int_{D_v} f\, d\mu_{v,\lambda} := c_{v,\lambda} \int_{D_v} f(z)\, h_v(z,z)^{\lambda - p_v}\, dm_v(z), \tag{5.4}$$

where m_v is the Lebesgue measure on D_v and $c_{v,\lambda}$ is the normalization factor. Similarly, one defines a probability measure σ_v on the Shilov boundary S_v of D_v, via

$$\int_{S_v} f\, d\sigma_v := \int_{K_v} f(k(v'))\, dk,$$

where v' is any tripotent orthogonal to v and $K_v := \{k \in K; k(Z_\nu(v)) = Z_\nu(v)\}$, $\nu = 0, 1/2, 1$, so that $K_v(v') = S_v$. The combination of $\mu_{v,\lambda}$ and σ_ℓ yields K-invariant probability measures $\mu_{\ell,\lambda}$ on $\partial_\ell D$, $1 \leq \ell \leq r-1$, $\lambda > p - \ell\, a - 1$, via

$$\int_{\partial_\ell D} f d\mu_{\ell,\lambda} := \int_{S_\ell} \left(\int_{D_v} f_v(z)\, d\mu_{v,\lambda}(z) \right)\, d\sigma_\ell(v).$$

Next, consider the "sphere bundle" B_ℓ, $1 \leq \ell \leq r$, whose base is S_ℓ and the fiber at each $v \in S_\ell$ is $v + S_v$ (where $S_v := \partial_{r-\ell} D_v$ is the Shilov boundary of D_v). The group K acts on B_ℓ naturally, and this action is transitive. The combination of the measures σ_v, $v \in S_\ell$ and σ_ℓ yields K-invariant probability measures ν_ℓ on B_ℓ via

$$\int_{B_\ell} f\, d\nu_\ell := \int_{S_\ell} \left(\int_{S_v} f(v+\xi)\, d\sigma_v(\xi) \right) d\sigma_\ell(v).$$

For $v \in S_\ell$, consider the symmetric cone Ω_v in $Z_0(v)$, and let $\Delta_1^{(v)}, \Delta_2^{(v)}, \ldots, \Delta_{r-\ell}^{(v)}$ be the canonical generators of the ring $\text{Diff}(\Omega_v)^{G(\Omega_v)}$ as in section 4. We also denote

$$\Delta_0^{(v)} = I,\ \boldsymbol{\Delta}^{(v)} := (\Delta_1^{(v)}, \Delta_2^{(v)}, \ldots, \Delta_{r-\ell}^{(v)}), \text{ and } \lambda_j = (j-1)\frac{a}{2},\ 0 \leq j \leq r.$$

CONJECTURE: *For every $2 \leq j \leq r$ and every $\lambda > \lambda_{j-1}$ there exists a positive function $p_{j,\lambda} \in C^\infty([0,\infty)^{j-1})$, so that the inner product $\langle \cdot, \cdot \rangle_{\lambda_j} = \langle \cdot, \cdot \rangle_{\lambda_j, 0}$ is given by*

$$\langle f, g \rangle_{\lambda_j} = \int_{S_{r-j+1}} \langle p_{j,\lambda}(\boldsymbol{\Delta}^{(v)}) f_v, g_v \rangle_{\mathcal{H}_\lambda(D_v)}\, d\sigma_{r-j+1}(v). \tag{5.5}$$

Moreover, if $\lambda = \lambda_{j-1} + 1 = dim(D_v)/rank(D_v)$, then $p_j := p_{j,\lambda}$ is a polynomial with positive coefficients.

If λ is chosen appropriately then (5.5) becomes an integral formula for $\langle f, g \rangle_{\lambda_j}$. For instance, if $\lambda = \lambda_{j-1} + 1$ in (5.5), then we have $\mathcal{H}_\lambda(D_v) = H^2(S_v)$, and (5.5) becomes

$$\langle f, g \rangle_{\lambda_j} = \int_{S_{r-j+1}} \left(\int_{S_v} (p_{j,\lambda}(\boldsymbol{\Delta}^{(v)}) f_v)(\xi)\, \overline{g_v(\xi)} d\sigma_v(\xi) \right) d\sigma_{r-j+1}(v). \tag{5.6}$$

Also, if $\lambda > (j-2)a + 1$ in (5.5) then $\mathcal{H}_\lambda(D_v) = L_a^2(D_v, \mu_{v,\lambda})$, and (5.5) becomes

$$\langle f, g \rangle_{\lambda_j} = \int_{S_{r-j+1}} \left(\int_{D_v} (p_j(\boldsymbol{\Delta}^{(v)}) f_v)(z)\, \overline{g_v(z)}\, d\mu_{v,\lambda}(z) \right)\, d\sigma_{r-j+1}(v). \tag{5.7}$$

Note that the integral in (5.7) can be expressed as an integral on $\partial_{r-j+1} D$ with respect to $d\mu_{r-j+1,\lambda}$. Similarly, (5.6) is an integral on B_{r-j+1} with respect to ν_{r-j+1}.

INTEGRAL FORMULAS FOR $\langle f,g\rangle_{a/2}$ VIA INTEGRATION ON $\partial_{r-1}D$

In what follows we shall establish (5.5) for $j=2$ (i.e. $\lambda_2=\frac{a}{2}$) in two important special cases, namely for Cartan domains of type I and IV. Our method suggests an approach for the general case. For $j=2$ (5.5) becomes

$$\langle f,g\rangle_{\frac{a}{2}}=\int_{S_{r-1}} d\sigma_{r-1}(v)\langle p_\lambda(R^{(v)})f_v,g_v\rangle_{\mathcal{H}_\lambda(\mathbf{D})}, \tag{5.8}$$

where $p_\lambda(x)=p_{2,\lambda}(x)\in C^\infty([0,\infty))$ is a positive function, $\Delta_1^{(v)}=R^{(v)}$, where $R^{(v)}$ is the localized radial derivative (i.e. the radial derivative in $Z_0(v)$), and $D_v\equiv\mathbf{D}=\{z\in\mathbf{C};|z|<1\}$. We will show that in our two cases

$$p_\lambda(x)=\frac{\Gamma(x+\lambda)}{\Gamma(\lambda)\Gamma(x+1)}\ q(x),$$

where $q(x)$ is a polynomial with positive rational coefficients. In particular, for $\lambda=1,2,\ldots,$ $p_\lambda(x)$ itself is a polynomial with positive rational coefficients. If λ is chosen appropriately, then (5.8) becomes an integral formula analogous to (5.6) or (5.7). For $\lambda=1$, (5.8) becomes

$$\langle f,g\rangle_{\frac{a}{2}}=\int_{S_{r-1}} d\sigma_{r-1}(v)\langle p_1(R^{(v)})f_v,g_v\rangle_{H^2(\mathbf{T})}, \tag{5.9}$$

and for $\lambda>1$, (5.8) becomes

$$\langle f,g\rangle_{\frac{a}{2}}=\int_{S_{r-1}} d\sigma_{r-1}(v)\langle p_\lambda(R^{(v)})f_v,g_v\rangle_{L^2(\mathbf{D},\mu_\lambda)}. \tag{5.10}$$

LEMMA 5.1 *The right hand side of (5.5) is K-invariant. Consequently, the right hand sides of (5.6), (5.7), (5.8), (5.9), and (5.10) are K-invariant.*

PROOF: Let $\ell=r-j+1$, and note that for each fixed smooth function f the maps $S_\ell\ni v\mapsto\Delta_i^{(v)}(f_v)$, $1\le i\le j-1$, are K-invariant, in the sense that

$$\Delta_i^{(k(v))}(f_{k(v)})\circ k=\Delta_i^{(v)}((f\circ k)_v),\quad \forall k\in K,\ \ \forall v\in S_\ell.$$

From this it follows that if $v_\ell\in S_\ell$ is any fixed element, then

$$\begin{aligned}&\int_{S_\ell}\langle p_{j,\lambda}(\mathbf{\Delta}^{(v)})f_v,g_v\rangle_{\mathcal{H}_\lambda(D_v)}d\sigma_\ell(v)\\ =\ &\int_K\langle p_{j,\lambda}(\mathbf{\Delta}^{(v_\ell)})(f\circ k)_{v_\ell},(g\circ k)_{v_\ell}\rangle_{\mathcal{H}_\lambda(D_{v_\ell})}dk.\end{aligned}$$

The K-invariance of the right hand side of (5.5) follows from the invariance of the Haar measure dk. ∎

Since $\mathcal{M}_0^{(\frac{a}{2})}=\sum_{m=0}^\infty P_{(m,0,0,\ldots)}$ and

$$\langle f,g\rangle_{\frac{a}{2}}=\sum_{\mathbf{m}=(m,0,\ldots,0),0\le m<\infty}\frac{\langle f_\mathbf{m},g_\mathbf{m}\rangle_\mathcal{F}}{(\frac{a}{2})_\mathbf{m}},$$

in order to establish (5.8) it is enough, by the K-invariance of both sides, to find positive functions $p_\lambda(x) \in C^\infty([0,\infty))$ so that (5.8) holds for the functions $f(z) = g(z) = N_1^m(z)$, $m \geq 0$. This is equivalent to

$$\int_{S_{r-1}} d\sigma_{r-1}(v) \langle p_\lambda(R^{(v)})(N_1^m)_v, (N_1^m)_v\rangle_{H_\lambda(\mathbf{D})} = \frac{m!}{(\frac{a}{2})_m}. \tag{5.11}$$

Fix a frame $e_1, e_2, \ldots, e_r$ in Z. Then $N_1(z) = (z, e_1)$, where $(\cdot,\cdot)$ is the unique K-invariant inner product on Z for which $(v,v) = 1$ for every minimal tripotent v. Let $e' := e_2 + e_3 + \ldots + e_r$. Then for $z = k(\xi e_1 + e')$ with $k \in K$ and $\xi \in \mathbf{T}$, we have

$$N_1^m(z) = (\xi k(e_1) + k(e'), e_1)^m = \sum_{\ell=0}^m \binom{m}{\ell} (k(e_1), e_1)^\ell \, (k(e'), e_1)^{m-\ell} \, \xi^\ell.$$

Thus, for $v = k(e')$, $m \geq 0$ and any continuous function f we have

$$(f(R^{(v)})N_1^m)(z) = \sum_{\ell=0}^m \binom{m}{\ell} (k(e_1), e_1)^\ell \, (k(e'), e_1)^{m-\ell} \, f(\ell) \, \xi^\ell.$$

Let us define

$$J_{m,\ell} := \int_K |(k(e_1), e_1)|^{2\ell} \, |(k(e'), e_1)|^{2(m-\ell)} \, dk, \quad 0 \leq \ell \leq m < \infty. \tag{5.12}$$

It follows that the function p_λ should satisfy

$$\int_{S_{r-1}} d\sigma_{r-1}(v) \, \langle p_\lambda(R^{(v)})(N_1^m)_v, (N_1^m)_v\rangle_{\mathcal{H}_\lambda(\mathbf{D})} = \sum_{\ell=0}^m J_{m,\ell} \binom{m}{\ell}^2 \frac{\ell!}{(\lambda)_\ell} \, p_\lambda(\ell).$$

Thus (5.11) becomes

$$\sum_{\ell=0}^m J_{m,\ell} \binom{m}{\ell}^2 q_\ell = \frac{m!}{(\frac{a}{2})_m}, \quad m = 0, 1, 2, \ldots, \tag{5.13}$$

where the numbers

$$q_\ell := \frac{\ell!}{(\lambda)_\ell} \, p_\lambda(\ell), \quad \ell = 0, 1, 2, \ldots \tag{5.14}$$

do not depend on λ. The infinite system of linear equations (5.13) in the unknowns $\{q_\ell\}_{\ell=0}^\infty$ corresponds to the lower triangular matrix $A = (a_{m,\ell})_{m,\ell=0}^\infty$, where $a_{m,\ell} = J_{m,\ell} \binom{m}{\ell}^2$ for $m \geq \ell$, and $a_{m,\ell} = 0$ for $m < \ell$. Since $a_{m,m} > 0$ for $m = 0, 1, 2, \ldots$, there exists a unique solution $\{q_\ell\}_{\ell=0}^\infty$ to (5.13). There are many smooth functions which interpolate the values $\{q_\ell\}_{\ell=0}^\infty$. We will show that $q_\ell > 0$ for every $\ell \geq 0$, and prove that $\{q_\ell\}_{\ell=0}^\infty$ can be interpolated by a polynomial of degree $r-1$ with positive coefficients. For Cartan domains of type I and IV, we will solve the system (5.13) by calculating explicitly the numbers $J_{m,\ell}$ and applying powers of the difference operator

$$\delta(f)(t) := f(t) - f(t-1), \quad t \in \mathbf{R}.$$

If f is defined only on $[0,\infty)$ then we define $\delta(f) := \delta(F)$, where $F(t) := f(t)$ for $0 \le t$ and $F(t) = 0$ for $0 > t$. Similarly, δ can be defined on two-sided sequences (i.e. on functions on $\mathbf{Z}$) or on sequences (i.e. functions on $\mathbf{N}$). The powers of δ are defined inductively by $\delta^{n+1} := \delta \circ \delta^n$.

CASE 1: CARTAN DOMAINS OF TYPE I. Let $D = D(I_{r,r}) := \{z \in M_{r,r}(\mathbf{C}); \|z\| < 1\}$. The rank of D is r, the dimension is $d = r^2$, the genus is $p = 2r$, and the characteristic multiplicity is $a = 2$. To every $k \in K$ there correspond $u, w \in U(r)$ (the *unitary group*) so that $det(u) = det(w)$, and

$$k(z) = uzw^*, \quad z \in D. \tag{5.15}$$

Thus $\int_K f(k(z))\, dk = \int_{U(r)} \int_{U(r)} f(uzw^*)\, du\, dw$, where dk is the Haar measure of K. Choose the canonical frame of matrix units $e_j := e_{j,j}$, $j = 1, 2, \ldots, r$, and denote $e = \sum_{j=1}^r e_j$ and $e' := e - e_1 = \sum_{j=2}^r e_j$.

PROPOSITION 5.1 *Let $D = D(I_{r,r})$. Then for every integers m, ℓ with $0 \le \ell \le m < \infty$, we have*

$$J_{m,\ell} = \frac{(r-1)\,(\ell\,!)^2\,(m-\ell)!\,(m-\ell+r-2)!}{(r)_m\,(m+r-1)!}. \tag{5.16}$$

PROOF: Let $k \in K$ be given by (5.15). Then $(k(e_1), e_1) = u_{1,1}\overline{w_{1,1}}$ and $(k(e'), e_1) = \sum_{j=2}^r u_{1,j}\overline{w_{1,j}}$. Thus, for $0 \le \ell \le m < \infty$,

$$J_{m,\ell} = \int_{U(r)} \int_{U(r)} |u_{1,1}|^{2\ell} |w_{1,1}|^{2\ell} |\sum_{j=2}^r u_{1,j}\overline{w_{1,j}}|^{2(m-\ell)}\, du\, dw.$$

This integral can be written as an integral on the product of the unit spheres $\partial\mathbf{B}_r \subset \mathbf{C}^r$ with respect to the $U(r)$-invariant probability measure σ:

$$J_{m,\ell} = \int_{\partial\mathbf{B}_r} \int_{\partial\mathbf{B}_r} |\xi_1^\ell|^2 |\eta_1^\ell|^2 |(\xi', \eta')|^{2(m-\ell)}\, d\sigma(\xi)\, d\sigma(\eta),$$

where $\xi' := (\xi_2, \ldots, \xi_r)$ and $\eta' := (\eta_2, \ldots, \eta_r)$. Now, by the $U(r)$-invariance,

$$\begin{aligned}
&\int_{\partial\mathbf{B}_r} |\xi_1^\ell|^2 |(\xi', \eta')|^{2(m-\ell)}\, d\sigma(\xi) \\
&\quad = \|\eta'\|^{2(m-\ell)} \int_{\partial\mathbf{B}_r} |\xi_1^\ell|^2 |\xi_2^{m-\ell}|^2\, d\sigma(\xi) \\
&\quad = \|\eta'\|^{2(m-\ell)} \|\xi_1^\ell \xi_2^{m-\ell}\|^2_{\mathcal{H}_r(D)} = \|\eta'\|^{2(m-\ell)} \frac{\ell!(m-\ell)!}{(r)_m}.
\end{aligned}$$

It follows by using [Ru], 1.4.5, that

$$\begin{aligned}
J_{m,\ell} &= \frac{\ell!(m-\ell)!}{(r)_m} \int_{\partial\mathbf{B}_r} |\eta_1^\ell|^2 (1 - |\eta_1|^2)^{m-\ell}\, d\sigma(\eta) \\
&= \frac{\ell!(m-\ell)!}{(r)_m} (r-1) \int_0^1 t^\ell (1-t)^{m-\ell+r-2}\, dt
\end{aligned}$$

$$= \frac{\ell!(m-\ell)!}{(r)_m}(r-1)B(\ell+1, m-\ell+r-1)$$

$$= \frac{(r-1)(\ell!)^2(m-\ell)!(m-\ell+r-2)!}{(r)_m(m+r-1)!}. \qquad \blacksquare$$

Corollary 5.1 *For $D = D(I_{r,r})$ the system of equations (5.13) is equivalent to the system*

$$\sum_{\ell=0}^{m} \frac{(m-\ell+r-2)!}{(m-\ell)!}\, q_\ell = (r-2)! \binom{m+r-1}{r-1}^2, \quad m = 0, 1, 2, \ldots. \tag{5.17}$$

Proposition 5.2 *For every $r \geq 2$ there exists a polynomial $q(x) = q_r(x)$ of degree $r-1$ with positive rational coefficients, so that $q(\ell) = q_\ell$ for $\ell = 0, 1, 2, \ldots$, where $\{q_\ell\}_{\ell=0}^{\infty}$ is the unique solution of (5.17).*

For small values of r it is easy to solve (5.17) explicitly by applying powers of δ. Thus,

$$q_2(x) = 2x+1, \quad q_3(x) = 3x^2+3x+1, \quad \text{and} \quad q_4(x) = \frac{1}{3}(10x^3+15x^2+11x+3).$$

The proof in the general case requires more preparation. Define

$$f_n(x) := (x+1)_n = \prod_{j=1}^{n}(x+j), \quad n \geq 1, \text{ and } g_n(x) := \prod_{j=0}^{n}(x+j)^2, \quad n \geq 0. \tag{5.18}$$

Then $g_n(x+1) = f_{n+1}(x)^2$, and

$$(\delta^k f_n)(x) = n(n-1)\cdots(n-k+1)\, f_{n-k}(x), \quad k \geq 0, \tag{5.19}$$

where δ is defined by $\delta(f)(x) := f(x) - f(x-1)$. Indeed, (5.19) is trivial for $k = 0$. For $k = 1$ and all n we have

$$\delta(f_n)(x) = \prod_{j=1}^{n}(x+j) - \prod_{j=1}^{n}(x+j-1) = \prod_{j=1}^{n-1}(x+j)\,(x+n-x) = n\, f_{n-1}(x).$$

Assuming (5.19) for k, let $n > k$ and compute $\delta^{k+1}(f_n)(x) = n(n-1)\cdots(n-k+1)\,\delta(f_{n-k})(x) = n(n-1)\cdots(n-k+1)(n-k)f_{n-k-1}(x)$. This establishes (5.19).

Next, define an operator σ, analogous to δ, via

$$(\sigma f)(x) := f(x) + f(x-1), \quad x \in \mathbf{R}.$$

Clearly, $\delta\sigma = \sigma\delta$, and both σ and δ commute with all the translation operators

$$(\tau_c f)(x) := f(x+c).$$

Denote by $\mathcal{P}_+$ the set of polynomials in one variable with non-negative coefficients.

Lemma 5.2 *Let $f(x)$ be a polynomial and let $n, m \in \mathbf{N}$. If $\delta^n f \in \mathcal{P}_+$, then $\delta^{n+j}\tau_{m/2} f \in \mathcal{P}_+$ for every integer $0 \leq j \leq m$.*

Proof: Since δ commutes with translations, we may assume that $n = 0$ and $m = 1$. It is therefore enough to check that $\delta\tau_{1/2} x^k \in \mathcal{P}_+$ for every $k \in \mathbf{N}$. This follows from the binomial expansion:

$$\delta\tau_{1/2} x^k = (x + \frac{1}{2})^k - (x - \frac{1}{2})^k = \sum_{j=0}^{[\frac{k-1}{2}]} \binom{k}{2j+1} 2^{-2j} x^{k-2j-1}. \qquad \blacksquare$$

Lemma 5.3 *Let $f(x)$ be a polynomial and let $n \in \mathbf{N}$. Assume that $\delta^j\sigma^{n-j} f \in \mathcal{P}_+$ for every $0 \leq j \leq n$. Then $\delta^j\sigma^{n-j}\left((x+c)^k f(x)\right) \in \mathcal{P}_+$ for every $k \in \mathbf{N}$, $c \geq \frac{n}{2}$ and $0 \leq j \leq n$.*

Proof: Again, since δ and σ commute with translations, it is enough to assume that $k = 1$. We shall prove the assertion by induction on n. The case $n = 0$ is trivial since $\mathcal{P}_+$ is closed under sums and products. Assume that $n > 0$ and that the assertion holds for $n - 1$. A computation yields

$$\delta\left((x + \frac{n}{2}) f(x)\right) = (x + \frac{n-1}{2})(\delta f)(x) + \frac{1}{2}(\sigma f)(x) \tag{5.20}$$

and

$$\sigma\left((x + \frac{n}{2}) f(x)\right) = (x + \frac{n-1}{2})(\sigma f)(x) + \frac{1}{2}(\delta f)(x). \tag{5.21}$$

If $0 < j \leq n$ then using (5.20) we get

$$\delta^j\sigma^{n-j}\left((x + \frac{n}{2}) f(x)\right) = \delta^{j-1}\sigma^{(n-1)-(j-1)}\left((x + \frac{n-1}{2})(\delta f)(x) + \frac{1}{2}(\sigma f)(x)\right).$$

By assumption,

$$\delta^{j-1}\sigma^{(n-1)-(j-1)}\sigma f = \delta^{j-1}\sigma^{n-(j-1)} f \in \mathcal{P}_+, \quad \text{for } 0 < j \leq n.$$

Similarly,

$$\delta^{j-1}\sigma^{(n-1)-(j-1)}\delta f = \delta^j\sigma^{n-j} f \in \mathcal{P}_+ \quad \text{for } 0 < j \leq n.$$

Thus, by the induction hypothesis on $n - 1$,

$$\delta^{j-1}\sigma^{(n-1)-(j-1)}\left((x + \frac{n-1}{2})\delta f(x)\right) \in \mathcal{P}_+, \quad \text{for } 0 < j \leq n.$$

Next, using (5.21) we get

$$\sigma^n\left((x + \frac{n}{2}) f(x)\right) = \sigma^{n-1}\left((x + \frac{n-1}{2})\sigma f(x) + \frac{1}{2}\delta f(x)\right).$$

By assumption, $\sigma^{n-1}\delta f(x) \in \mathcal{P}_+$ and $\delta^\ell\sigma^{n-1-\ell}\sigma f(x) \in \mathcal{P}_+$ for $0 \leq \ell \leq n-1$. Thus, by the induction hypothesis, $\delta^\ell\sigma^{n-1-\ell}\left((x + \frac{n-1}{2})\sigma f(x)\right) \in \mathcal{P}_+$ for $0 \leq \ell \leq n-1$, and in particular $\sigma^{n-1}\left((x + \frac{n-1}{2})\sigma f(x)\right) \in \mathcal{P}_+$. It follows that $\sigma^n\left((x + \frac{n}{2}) f(x)\right) \in \mathcal{P}_+$. This completes the induction step. $\blacksquare$

LEMMA 5.4 *Let $g_n(x)$ be the polynomial defined by (5.18). Then $\delta^i\sigma^j g_n \in \mathcal{P}_+$ whenever $i+j \le n$.*

PROOF: We proceed by induction on n. The case $n=0$ is trivial, since $g_0(x) = x^2 \in \mathcal{P}_+$. Assume that $n>0$ and that $\delta^i\sigma^j g_{n-1} \in \mathcal{P}_+$ whenever $i+j \le n-1$. A computation yields

$$\delta g_n(x) = 2(n+1)(x+\frac{n-1}{2})\, g_{n-1}(x) \tag{5.22}$$

and

$$\sigma g_n(x) = 2\left((x+\frac{n-1}{2})^2 + (\frac{n+1}{2})^2\right) g_{n-1}(x). \tag{5.23}$$

Now assume $i+j \le n$. If $i>0$, (5.22) yields

$$\delta^i\sigma^j g_n(x) = \delta^{i-1}\sigma^j(\delta g_n(x)) = 2(n+1)\delta^{i-1}\sigma^j\left((x+\frac{n-1}{2})g_{n-1}(x)\right),$$

and by induction hypothesis and Lemma 5.3

$$\delta^{i-1}\sigma^j\left((x+\frac{n-1}{2})g_{n-1}(x)\right) \in \mathcal{P}_+,$$

so that $\delta^i\sigma^j g_n \in \mathcal{P}_+$. If $i=0$ and $0 \le j \le n$, then (5.23) implies

$$\sigma^j g_n(x) = \sigma^{j-1}(\sigma g_n(x)) = 2\sigma^{j-1}\left(\left((x+\frac{n-1}{2})^2 + (\frac{n+1}{2})^2\right) g_{n-1}(x)\right).$$

The polynomial $\sigma^{j-1}g_{n-1}$ belongs to $\mathcal{P}_+$ by the induction hypothesis. Also, the induction hypothesis ($\delta^i\sigma^{j-1}g_{n-1} \in \mathcal{P}_+$ whenever $i+j \le n$) and Lemma 5.3 imply that

$$\delta^i\sigma^{j-1}\left((x+\frac{n-1}{2})g_{n-1}(x)\right) \in \mathcal{P}_+ \quad \text{whenever } i+j \le n.$$

In particular, $\sigma^{j-1}\left((x+\frac{n-1}{2})g_{n-1}(x)\right) \in \mathcal{P}_+$. Hence $\sigma^j g_n \in \mathcal{P}_+$ $\forall\, 0 \le j \le n$. ∎

COROLLARY 5.2 (I) $\delta^j g_n \in \mathcal{P}_+$ *for all $j,n \in \mathbf{N}$ satisfying $0 \le j \le n$.*

(II) $\delta^j\left((x+\frac{m}{2})g_n(x)\right) \in \mathcal{P}_+$ *for all $j,n,m \in \mathbf{N}$ satisfying $0 \le j \le n+m$.*

(III) $\delta^j f_n(x)^2 \in \mathcal{P}_+$ *for all $j,n \in \mathbf{N}$ satisfying $0 \le j \le n+1$.*

PROOF: (i) is a special case of Lemma 5.4, and (ii) follows by (i) and Lemma 5.2. Since $f_n(x)^2 = g_{n-1}(x+1)$, (iii) follows from Lemma 5.2 with $m=2$. ∎

REMARK The result in part (iii) of Corollary 5.2 is best possible in the sense that $\delta^{n+2}(f_n^2)^2)$ need not be in $\mathcal{P}_+$. Indeed, $\delta^6(f_4^2)^2)$ is not in $\mathcal{P}_+$.

PROOF OF PROPOSITION 5.2: In terms of the polynomials (5.18), the system of equations (5.17) with unknowns q_ℓ has the form

$$\sum_{\ell=0}^{m} f_{r-2}(m-\ell)\, q_\ell = \frac{f_{r-1}(m)^2}{(r-1)\,(r-1)!}, \quad m \geq 0. \tag{5.24}$$

Applying powers of the operator δ with respect to the variable m and using (5.19), we get by induction on k that

$$\delta^k \left(\sum_{\ell=0}^{m} f_{r-2}(m-\ell)\, q_\ell \right) = (r-2)(r-3)\cdots(r-k-1) \sum_{\ell=0}^{m} f_{r-2-k}(m-\ell)\, q_\ell$$

for $0 \leq k \leq r-2$ (here $f_0(x) \equiv 1$). From this it follows that

$$\delta^{r-1} \left(\sum_{\ell=0}^{m} f_{r-2}(m-\ell)\, q_\ell \right) = (r-2)!\, q_m, \quad m \geq 0.$$

Applying δ^{r-1} to both sides of (5.24), Corollary 5.2 (iii) implies that there exists a polynomial $q(x)$ of degree $r-1$ with positive rational coefficients so that $q_m = q(m)$, $\forall m \geq 0$. ∎

THEOREM 5.1 *Let $D = D(I_{r,r})$. Then for every $f, g \in \mathcal{H}_{\frac{a}{2}}(D)$ and $\lambda > 0$ we have*

$$\langle f, g \rangle_{\frac{a}{2}} = \int_{S_{r-1}} d\sigma_{r-1}(v) \langle p_\lambda(R^{(v)}) f_v, g_v \rangle_{\mathcal{H}_\lambda(\mathbf{D})},$$

where $p_\lambda(x) := \Gamma(x+\lambda)\,\Gamma(\lambda)^{-1}\,\Gamma(x+1)^{-1}\, q(x)$, and $q(x)$ is the polynomial of degree $r-1$ with positive rational coefficients as in Proposition 5.2.

CASE 2: CARTAN DOMAINS OF TYPE IV. Let $D \subset \mathbf{C}^d$, $d \geq 3$, be the Cartan domain of rank $r = 2$ (see Examples 4.1 and 4.2), and fix a frame $\{e_1, e_2\}$. Since $a = d-2$, (5.13) becomes

$$\sum_{\ell=0}^{m} \binom{m}{\ell}^2 J_{m,\ell}\, q_\ell = \frac{m!}{(\frac{a}{2}-1)_m}, \quad m \geq 0, \tag{5.25}$$

where for $0 \leq \ell \leq m$

$$J_{m,\ell} = \int_K |(k(e_1), e_1)|^{2\ell} |(k(e_2), e_1)|^{2(m-\ell)}\, dk.$$

Without computing the numbers $J_{m,\ell}$ explicitly we show that

$$J_{m,\ell} = J_{m,m-\ell}, \quad 0 \leq \ell \leq m. \tag{5.26}$$

Indeed, let $k' \in K$ satisfy $k'(e_1) = e_2$ and $k'(e_2) = e_1$. Then, by invariance of the Haar measure dk,

$$\begin{aligned} J_{m,\ell} &= \int_K |(k(k'(e_1)), e_1)|^{2\ell} |(k(k'(e_2)), e_1)|^{2(m-\ell)}\, dk \\ &= \int_K |(k(e_2), e_1)|^{2\ell} |(k(e_1), e_1)|^{2(m-\ell)}\, dk = J_{m,m-\ell}. \end{aligned}$$

THEOREM 5.2 *The polynomial*

$$q(x) = \frac{4}{a}\,x + 1 = \frac{4}{d-2}\,x + 1$$

satisfies $q(\ell) = q_\ell$ *for every* $\ell \geq 0$, *where* $\{q_\ell\}_{\ell=0}^{\infty}$ *is the unique solution of (5.25). Therefore, for every* $\lambda > 0$ *and every* $f, g \in \mathcal{H}_{\frac{a}{2}}(D)$,

$$\langle f, g\rangle_{\frac{a}{2}} = \int_{S_1} \langle p_\lambda(R^{(v)}) f_v, g_v\rangle_{\mathcal{H}_\lambda(D_v)}\, d\sigma_1(v),$$

where the functions p_λ, $0 < \lambda < \infty$, *are given by*

$$p_\lambda(x) = \frac{\Gamma(x+\lambda)}{\Gamma(\lambda)\,\Gamma(x+1)}\,(\frac{4}{a}\,x + 1). \tag{5.27}$$

In particular, for $\lambda = 1, 2, \ldots$ p_λ *is a polynomial of degree* λ *with positive rational coefficients.*

PROOF: We claim first that

$$\sum_{\ell=0}^{m} \binom{m}{\ell}^2 J_{m,\ell} = \frac{m!}{(\frac{d}{2})_m}, \quad m \geq 0. \tag{5.28}$$

Indeed, it is clear that

$$\sum_{\ell=0}^{m} \binom{m}{\ell}^2 J_{m,\ell} = \int_K \left(\int_{\mathbf{T}} |(k(e^{it}e_1 + e_2), e_1)^m|^2\, \frac{dt}{2\pi}\right) dk.$$

Interchanging the order of integration and using the transitivity of K on the frames, we get

$$\sum_{\ell=0}^{m} \binom{m}{\ell}^2 J_{m,\ell} = \int_K |(k(e), e_1)^m|^2\, dk = \|N_1^m\|^2_{H^2(D)} = \frac{m!}{(\frac{d}{2})_m}, \quad m \geq 0,$$

by using the well-known fact that $\|(\cdot, z)^m\|^2_{\mathcal{F}} = m!(z,z)^m$ for every $z \in Z$ and $m \geq 0$. Using (5.26) and (5.28) we see that

$$\begin{aligned}
\sum_{\ell=0}^{m} \ell \binom{m}{\ell}^2 J_{m,\ell} &= \sum_{\ell=0}^{m} (m-\ell) \binom{m}{m-\ell}^2 J_{m,m-\ell} \\
&= \sum_{\ell=0}^{m} (m-\ell) \binom{m}{\ell}^2 J_{m,\ell} = \frac{m \cdot m!}{(\frac{d}{2})_m} - \sum_{\ell=0}^{m} \ell \binom{m}{\ell}^2 J_{m,\ell}.
\end{aligned}$$

Thus

$$\sum_{\ell=0}^{m} \ell \binom{m}{\ell}^2 J_{m,\ell} = \frac{m \cdot m!}{2(\frac{d}{2})_m}, \quad m \geq 0. \tag{5.29}$$

Combining (5.28) and (5.29), and using the fact that $(\frac{d}{2})_m = (\frac{a}{2})_m\ \frac{(\frac{a}{2}+m)}{\frac{a}{2}}$, we get for $m \geq 0$

$$\sum_{\ell=0}^{m} \binom{m}{\ell}^2 J_{m,\ell}\,(\frac{4}{a}\,\ell + 1) = \frac{4}{a}\,\frac{m \cdot m!}{2(\frac{d}{2})_m} + \frac{m!}{(\frac{d}{2})_m} = \frac{m!}{(\frac{a}{2})_m}.$$

In view of (5.14), this completes the proof. ∎

—large The computation of $\langle f,g\rangle_{p-1}$ by integration on $\partial_1 D$

We conclude this section with the derivation of a formula for $\langle f,g\rangle_{p-1}$ via integration on $\partial_1 D$.

Proposition 5.3 *Let $F\in C(\overline{D})$. Then*

$$\lim_{\lambda\downarrow p-1}\int_D F(z)\,d\mu_\lambda(z)=\int_{S_1}\left(\int_{D_v}F_v(w)\,d\mu_{v,p-1}(w)\right)d\sigma_1(v), \tag{5.30}$$

where the measures $\mu_{v,p-1}$ are defined by (5.4).

Proof: Using (1.13) and (1.14) as well as (1.22), (1.23), and (1.9), we can write

$$\begin{aligned}\int_D F(z)\,d\mu_\lambda(z) &= c_0\,c(\lambda)\int_{\mathbf{R}^r_+}F^\#(t)w(t)^a\prod_{j=1}^r(1-t_j)^a\,dt\\ &= c_0\,c(\lambda)\int_0^1\psi(t_1)\,(1-t_1)^{\lambda-p}\,dt_1,\end{aligned}$$

where

$$\psi(t_1):=\int_{[0,t_1)^{r-1}_+}F^\#(t_1,t')\prod_{1\le i<j\le r}(t_i-t_j)^a\prod_{j=2}^r(1-t_j)^{\lambda-p}\,dt',$$

and $c(\lambda)=c_D(\lambda)$ is given by (1.22). Here $t':=(t_2,t_3,\dots,t_r)$, $dt':=dt_2\,dt_3\dots dt_r$, and $[0,t_1)^{r-1}_+:=\{t'\in\mathbf{R}^{r-1};t_2>t_3>\dots>t_r>0\}$. Since $\psi\in C([0,1])$, we have $\lim_{\epsilon\downarrow 0}\left(\epsilon\int_0^1\psi(t)(1-t)^{\epsilon-1}\,dt\right)=\psi(1)$. Since $\lim_{\lambda\downarrow p-1}\ \Gamma(\lambda-p+1)\,(\lambda-p+1)=1$ and $c(p-1)=0$, we get

$$\begin{aligned}\lim_{\lambda\downarrow p-1}\int_D F(z)\,d\mu_\lambda(z) &= b\,\psi(1)\\ &= b\int_{[0,1)^{r-1}_+}F^\#(1,t')\prod_{2\le i<j\le r}(t_i-t_j)^a\prod_{j=2}^r(1-t_j)^{a-1}\,dt',\end{aligned}$$

where $b:=c_0\,c'(p-1)$. Using the definitions (5.1), (5.3) and the fact that for $v\in S_1$ the genus of D_v is $p-a$, we have (with the obvious meaning of the constants)

$$\begin{aligned}&\int_{S_1}\left(\int_{D_v}F_v(w)\,d\mu_{v,p-1}(w)\right)d\sigma_1(v)\\ &=c_{D_{e_1}}(p-1)\int_K\left(\int_{D_{e_1}}F_{k(e_1)}(k(\xi))\;h(k(\xi),k(\xi))^{a-1}\,dm(k(\xi))\right)dk\end{aligned}$$

$$= c_{D_{e_1}}(p-1)\, c_0(D_{e_1})$$

$$\times \int_K \left(\int_{[0,1)_+^{r-1}} \left(\int_{K_{e_1}} F(k(e_1 + k'(\sum_{j=2}^r t_j^{\frac{1}{2}} e_j\, dk')) \right) w(t')^a \prod_{j=2}^r (1-t_j)^{a-1}\, dt' \right) dk,$$

where $K_{e_1} := \{k \in K; k(e_1) = e_1\}$ and $w(t') := \prod_{2 \le i < j \le r}(t_i - t_j)^a$. Interchanging the order of integration, and using the fact that $k'(e_1) = e_1$ and the invariance of the Haar measure dk, we see that the last expression is equal to

$$c_{D_{e_1}}(p-1)\, c_0(D_{e_1}) \int_{[0,1)_+^{r-1}} F^{\#}(1,t')\, w(t')^a \prod_{j=2}^r (1-t_j)^{a-1}\, dt'.$$

Comparing the computations for the left and right hand sides of (5.30), we see they are proportional. Taking $F(z) \equiv 1$, the proportionality constant is 1. ∎

COROLLARY 5.3 *The constant $c_0 = c_0(D)$ in the formula (1.12) is*

$$c_0(D) = \frac{\pi^d\, \Gamma(\frac{a}{2})^{r-2}}{(\prod_{\ell=1}^{r-1} \ell\, \frac{a}{2})\, \Gamma(r\, \frac{a}{2}) \prod_{\ell=2}^{r-1} \Gamma(\ell\, \frac{a}{2})^2}.$$

PROOF: Define $v_r = 0$, $v_\ell := e_1 + \ldots + e_{r-\ell}$, $\ell = 1, 2, \ldots, r-1$, and $\gamma_\ell := c_0(D_{v_\ell})$. Then the above proof (with r replaced by ℓ) yields

$$\frac{\gamma_\ell}{\gamma_{\ell-1}} = \frac{c_{D_{v_{\ell+1}}}((\ell-1)a+1)}{c'_{D_{v_\ell}}((\ell-1)a+1)} = \frac{\pi^{(\ell-1)a+1}\, \Gamma(\frac{a}{2})}{\Gamma((\ell-1)\frac{a}{2}+1)\, \Gamma(\frac{ra}{2})}$$

for $\ell = 2, 3, \ldots, r$. Therefore, using the easily proved fact that $\gamma_1 = \pi$, we get

$$\begin{aligned} c_0(D) &= \gamma_r = \frac{\gamma_r}{\gamma_{r-1}} \frac{\gamma_{r-1}}{\gamma_{r-2}} \cdots \frac{\gamma_2}{\gamma_1} \gamma_1 \\ &= \pi \prod_{\ell=2}^r \frac{\pi^{(\ell-1)a+1}\, \Gamma(\frac{a}{2})}{\Gamma((\ell-1)\frac{a}{2}+1)\, \Gamma(\frac{ra}{2})} = \frac{\pi^d\, \Gamma(\frac{a}{2})^{r-2}}{(\prod_{\ell=1}^{r-1} \ell\, \frac{a}{2})\, \Gamma(r\, \frac{a}{2}) \prod_{\ell=2}^{r-1} \Gamma(\ell\, \frac{a}{2})^2}. \end{aligned}$$ ∎

Proposition 5.3 allows the computation of the inner products $\langle f, g \rangle_{p-1}$ by integrating over the boundary orbit $\partial_1(D) = G(e_1)$ of G.

THEOREM 5.3 *Let $f, g \in \mathcal{H}_{p-1}$. Then*

$$\langle f, g \rangle_{p-1} = \int_{S_1} \left(\int_{D_v} f_v(w)\, \overline{g_v(w)}\, d\mu_{v,p-1}(w) \right) d\sigma_1(v), \tag{5.31}$$

PROOF: It is enough to establish (5.31) for polynomials f and g, and this case follows from Proposition 5.3 with $F(z) = f(z)\, \overline{g(z)}$. ∎

6 Integral formulas in the context of the associated Siegel domain

In what follows we shall use the fact [FK2] that D is holomorphically equivalent to the tube domain

$$T(\Omega) := X + i\Omega$$

via the *Cayley transform* $c : D \to T(\Omega)$, defined by $c(z) := i(e+z)(e-z)^{-1}$. For $\lambda \in W(D)$ the operator $V^{(\lambda)}f := (f\circ c^{-1})(Jc^{-1})^{\lambda/p}$ maps the space $\mathcal{H}_\lambda = \mathcal{H}_\lambda(D)$ isometrically onto a Hilbert space of analytic functions on $T(\Omega)$, denoted by $\mathcal{H}_\lambda(T(\Omega))$. We will denote $\langle f,g\rangle_{\mathcal{H}_\lambda(T(\Omega))}$ simply by $\langle f,g\rangle_\lambda$. It is known that the reproducing kernel of $\mathcal{H}_\lambda(T(\Omega))$ is

$$K_\lambda(z,w) = \left(N(\frac{z-w^*}{i})\right)^{-\lambda}, \quad z,w \in T(\Omega). \tag{6.1}$$

Recall that for $\lambda > p-1$ we have $\mathcal{H}_\lambda(D) = L^2_a(D,\mu_\lambda)$, where μ_λ is the measure on D defined via (1.23). Using the facts that $h(c^{-1}(w),c^{-1}(w)) = 4^r|N(w+ie)|^{-2}N(v)$ and $J(c^{-1})(w) = (2i)^d N(w+ie)^{-p}$, $\quad \forall w \in T(\Omega)$, we get by a change of variables that

$$\mathcal{H}_\lambda(T(\Omega)) = L^2_a(T(\Omega),\nu_\lambda) = L^2(T(\Omega),\nu_\lambda) \cap \{\text{analytic functions}\},$$

where

$$d\nu_\lambda(z) := c(\lambda)dx \; N(2y)^{\lambda-p}\,dy, \quad z = x+iy,\ x\in X,\ y\in\Omega, \tag{6.2}$$

and $c(\lambda)$ is defined by (1.22). In this case $V^{(\lambda)}$ extends to an isometry of $L^2(D,\mu_\lambda)$ onto $L^2(T(\Omega),\nu_\lambda)$.

In this section we obtain integral formulas for the invariant inner products in the spaces $\mathcal{H}_\lambda(T(\Omega))$. Using the isometry $V^{(\lambda)} : \mathcal{H}_\lambda(D) \to \mathcal{H}_\lambda(T(\Omega))$ one obtains integral formulas for the inner products in the spaces $\mathcal{H}_\lambda(D)$. Our results are essentially implicitly contained in [VR], where the authors determine the Wallach set for Siegel domains of type II, using Lie and Fourier theoretical methods. The Jordan-theoretical formalism allows us to formulate our results in a simpler way, avoiding the Lie-theoretical details. Since the Fourier-theoretical arguments in our proofs are contained in[VR], we omit all proofs.

For $\lambda > (r-1)\frac{a}{2}$ consider the measure σ_λ on Ω defined by $d\sigma_\lambda(v) := \beta_\lambda\, N(v)^{\frac{d}{r}-\lambda}\,dv$ where $\beta_\lambda := (2\pi)^{-2d}\Gamma_\Omega(\lambda)$.

Proposition 6.1 *Let $\lambda > (r-1)\frac{a}{2}$ and let f be a holomorphic function on $T(\Omega)$. Then the following conditions are equivalent:*

(i) $f \in \mathcal{H}_\lambda(T(\Omega))$;

(ii) *The boundary values $f(x) := \lim_{\Omega\ni y\to 0} f(x+iy)$ exist almost everywhere on X, and the Fourier transform $\hat{f}$ of $f(x)$ is supported in $\overline{\Omega}$ and belongs to $L^2(\Omega,\sigma_\lambda)$.*

Moreover, the map $f \mapsto \hat{f}$ is an isometry of $\mathcal{H}_\lambda(T(\Omega))$ onto $L^2_a(\Omega,\sigma_\lambda)$.

Proposition 6.1 yields the following result.

THEOREM 6.1 *Let* $\lambda > (r-1)\frac{a}{2}$ *and let* $f, g \in \mathcal{H}_\lambda(T(\Omega))$. *Then*

$$\langle f, g\rangle_{\mathcal{H}_\lambda(T(\Omega))} = \langle \hat{f}, \hat{g}\rangle_{L^2(\Omega,\sigma_\lambda)} = \frac{\Gamma_\Omega(\lambda)}{(2\pi)^{2d}} \int_\Omega \hat{f}(t)\overline{\hat{g}(t)}\, N(t)^{\frac{d}{r}-\lambda}\, dt.$$

The group $GL(\Omega) := \{\varphi \in GL(X); \varphi(\Omega) = \Omega\}$ acts transitively on Ω . It acts also on the boundary $\partial\Omega$, but this action is not transitive. The orbits of $GL(\Omega)$ on $\partial\Omega$ are exactly the r disjoint sets

$$\partial_k\Omega := GL(\Omega)(e_k) = \{x \in \overline{\Omega};\ \mathrm{rank}(x) = k\}, \quad k = 0, 1, \ldots, r-1,$$

where $\{c_1, \ldots, c_r\}$ is a frame of pairwise orthogonal primitive idempotents, $e_0 := 0$, and $e_k := \sum_{j=1}^k c_j$, $k = 1, 2, \ldots, r-1$. Consider the Peirce decomposition $X_\nu = X_\nu(e_k) = \{x \in X; e_k x = \nu x\}$, $\nu = 0, \frac{1}{2}, 1$. Let $\Omega(k)$ be the symmetric cone of $X_1(e_k)$, and let $\Gamma_{\Omega(k)}$ be the associated Gamma function. Let $GL(\Omega) = LN_\Omega A$ be the Iwasawa decomposition. Then $N_\Omega A(e_k) = \{x \in \partial_k\Omega; N_k(x) > 0\}$ is an open dense subset of $\partial_k\Omega$, and every $x \in N_\Omega A(e_k)$ has a Peirce decomposition of the form $x = x_1 + x_{\frac{1}{2}} + 2(e-e_k)(x_{\frac{1}{2}}(x_{\frac{1}{2}}x_1^{-1}))$ [La2]. Let us define a measure ν_k on $\partial_k\Omega$ with support $N_\Omega A(e_k)$ by

$$d\nu_k(x) := N_k(x_1)^{k\frac{a}{2}-\frac{d}{r}}\, dx_1\, dx_{\frac{1}{2}}. \tag{6.3}$$

It has the following fundamental properties (see[VR] and [La2]).

THEOREM 6.2 *Let* $1 \le k \le r-1$. *Then the measure* ν_k *satisfies*

$$\int_{N_\Omega A(e_k)} e^{-\langle y,x\rangle}\, d\nu_k(x) = \gamma_k\, N(y)^{-k\frac{a}{2}}, \quad \forall y \in \Omega, \tag{6.4}$$

where $\gamma_k := (2\pi)^{k(r-k)\frac{a}{2}}\,\Gamma_{\Omega(k)}(k\frac{a}{2})$, *and*

$$d\nu_k(\varphi(x)) = Det(\varphi)^{(k\frac{a}{2})/\frac{d}{r}}\, d\nu_k(x), \quad \forall \varphi \in GL(\Omega). \tag{6.5}$$

Since Ω is a set of uniqueness for analytic functions on $T(\Omega)$, (6.4) implies by analytic continuation

$$\int_{N_\Omega A(e_k)} e^{-\langle \frac{z-w^*}{i}, x\rangle}\, d\nu_k(x) = \gamma_k\, 2^{-k\frac{a}{2}} \left(N(\frac{z-w^*}{i})\right)^{-k\frac{a}{2}}, \quad \forall z, w \in T(\Omega).$$

Thus $\left(N(\frac{z-w^*}{i})\right)^{-k\frac{a}{2}}$ is positive definite, and so $k\frac{a}{2}$ is in the Wallach set $W(D) = W(T(\Omega))$.

By complexification, $GL(\Omega)$ is realized as a subgroup of $Aut(T(\Omega))$ which normalizes the translations $\tau_x(z) := z + x$, i.e.

$$\varphi\, \tau_x\, \varphi^{-1} = \tau_{\varphi(x)}, \quad \forall x \in X, \quad \forall \varphi \in GL(\Omega).$$

Let $G \subset Aut(T(\Omega))$ be the semi-direct product of X and $GL(\Omega)$. It acts transitively on $T(\Omega)$. Let $N \subset G$ be the semi-direct product of X and N_Ω. Then the Iwasawa decomposition of $Aut(T(\Omega))_0$ is KAN. For

$$\alpha_k = \frac{d}{r} + k\frac{a}{2}, \quad k = 0, 1, 2, \ldots, r-1$$

let $\mathcal{H}_{\alpha_k} = \mathcal{H}_{\alpha_k}(T(\Omega))$ be the Hilbert space of analytic functions on $T(\Omega)$ whose reproducing kernel is $K_{\alpha_k}(z,w) := \left(N(\frac{z-w^*}{i})\right)^{-\alpha_k}$. Note that $\alpha_{r-1} = p-1$ and for $k=0$ we have $\alpha_0 = \frac{d}{r}$ and $\nu_0 = \delta_0$, the Dirac measure at 0.

THEOREM 6.3 *For $k = 0, 1, \ldots, r-1$ $\mathcal{H}_{\alpha_k}(T(\Omega))$ consists of all analytic functions f on $T(\Omega)$ for which*

$$\|f\|^2_{\mathcal{H}_{\alpha_k}(T(\Omega))} := \beta_k \sup_{t\in\Omega} \int_{N_\Omega A(e_k)} \left(\int_X |f(x+i(y+t))|^2\, dx\right) d\nu_k(y) \tag{6.6}$$

is finite, where

$$\beta_k = \frac{\Gamma_\Omega(\alpha_k) 2^{rk\frac{a}{2}}}{\Gamma_{\Omega(k)}(k\frac{a}{2})} (2\pi)^{-(d+k(r-k)\frac{a}{2})}.$$

Moreover, for every $f, g \in \mathcal{H}_{\alpha_k}(T(\Omega))$,

$$\langle f, g\rangle_{\alpha_k} = \beta_k \lim_{\Omega\ni t\to 0} \int_{N_\Omega A(e_k)} \left(\int_X f(x+i(y+t))\, \overline{g(x+i(y+t))}\, dx\right) d\nu_k(y).$$

REFERENCES

[A1] J. Arazy, *A survey of invariant Hilbert spaces of analytic functions on bounded symmetric domains*, Contemp. Math. 185 (1995), 7-65.

[A2] J. Arazy, *Realization of the invariant inner products on the highest quotients of the composition series*, Arkiv Mat. 30 (1992), 1-24.

[A3] J. Arazy, *Integral formulas for the invariant inner products in spaces of analytic functions in the unit ball*, in Function Spaces, Lecture Notes in Pure and Applied Mathematics Vol.136, Marcel Dekker (1992), 9-23.

[A4] J. Arazy, *Boundedness and compactness of generalized Hankel operators on bounded symmetric domains*, J. Funct. Anal. 137 (1996), 97-151.

[AF] J. Arazy and S.D. Fisher, *Invariant Hilbert spaces of analytic functions on bounded symmetric domains*, In: Topics in Operator Theory, E.D. Hellinger Memorial Volume (Editors: L. de-Branges, I. Gohberg and J. Rovnyak), Operator Theory, Advances and Applications, 48 (1990), 67-91.

[Be] F. Berezin, *Quantization in complex symmetric spaces*, Math. USSR-Izv. 9 (1975), 341-379.

[FK1] J. Faraut and A. Korányi, *Function spaces and reproducing kernels on bounded symmetric domains*, J. Funct. Anal. 88 (1990), 64-89.

[FK2] J. Faraut and A. Korányi, *Analysis on Symmetric Cones*, Clarendon Press, Oxford (1994).

[Gi] S. Gindikin, *Analysis on homogeneous domains*, Russ. Math. Surv. 19 (1964), 1-89.

[Go] L. Garding, *The solution of Cauchy's problem for two totally hyperbolic linear differential equations by means of Riesz integrals*, Ann. Math. 48 (1947), 785-826.

[He] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric spaces*, Academic Press (1978).

[Hu] L.K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains*, Translations of Mathematical Monographs, Amer. Math. Soc. (1963).

[La1] M. Lassalle, *Noyau de Szegö, K-types et algèbres de Jordan*, C. R. Acad. Sci. Paris 303 (1986), 1-4.

[La2] M. Lassalle, *Algèbres de Jordan et ensemble de Wallach*, Invent. Math. 89 (1087), 375-393.

[Lo] O. Loos, *Bounded Symmetric Domains and Jordan Pairs*, University of California at Irvine (1977).

[O] B. Orsted, *Composition series for analytic continuations of holomorphic discrete series representations of $SU(n,n)$*, Trans. Amer. Math. Soc. 260 (1980), 563-573.

[Pel] M. Peloso, *Möbius invariant spaces on the unit ball*, Mich. Math. J. 39 (1992), 509-536.

[Ri] M. Riesz, *L'intégrale de Riemann-Liouville et le problème de Cauchy*, Acta Math. 81 (1949), 1-223.

[Ru] W. Rudin, *Function theory in the unit ball of* $\mathbf{C}^n$, Springer-Verlag, New York (1980).

[Sc] W. Schmid, *Die Randwerte holomorpher Funktionen auf hermitesch symmetrischen Räumen*, Invent. Math. 8 (1969), 61-80.

[T] A. Terras, Harmonic analysis on symmetric spaces and applications II, Springer-Verlag (1988).

[U1] H. Upmeier, *Jordan algebras and harmonic analysis on symmetric spaces*, Amer. J. Math. 108 (1986), 1-25.

[U2] H. Upmeier, *Jordan algebras in analysis, operator theory, and quantum mechanics*, CBMS Series in Math. 67, Amer. Math. Soc., (1987).

[VR] M. Vergne and H. Rossi, Analytic continuation of the holomorphic discrete series of a semi-simple Lie group, Acta Mathematica 136 (1976), 1-59.

[Y1] Z. Yan, *Invariant Differential Operators on Symmetric Cones*, preprint (1991).

[Y2] Z. Yan, *Invariant Differential Operators and Holomorphic Function Spaces*, preprint (1991).

[Y3] Z. Yan, *Duality and Differential Operators on the Bergman Space of Bounded Symmetric Domains*, J. Funct. Anal. 105 (1992), 171 - 186.

[Y4] Z. Yan, *Differential operators and function spaces*, Contemp. Math. 142 (1993), 121-142.

Jonathan Arazy
Department of Mathematics
University of Haifa
Haifa 31905, Israel
jarazy@mathcs2.haifa.ac.il

Harald Upmeier
Fachbereich Mathematik
Universität Marburg
35032 Marburg, Germany
upmeier@mathematik.uni-marburg.de

Higher Index Theorems and the Boundary Map in Cyclic Cohomology

Victor Nistor[1]

Received: June 23, 1997

Communicated by Joachim Cuntz

Abstract. We show that the Chern–Connes character induces a natural transformation from the six term exact sequence in (lower) algebraic K–Theory to the periodic cyclic homology exact sequence obtained by Cuntz and Quillen, and we argue that this amounts to a general "higher index theorem." In order to compute the boundary map of the periodic cyclic cohomology exact sequence, we show that it satisfies properties similar to the properties satisfied by the boundary map of the singular cohomology long exact sequence. As an application, we obtain a new proof of the Connes–Moscovici index theorem for coverings.

1991 Mathematics Subject Classification: (Primary) 19K56, (Secondary) 19D55, 46L80, 58G12.

Key Words: cyclic cohomology, algebraic K-theory, index morphism, etale groupoid, higher index theorem.

Contents

[1]Partially supported by NSF grant DMS 92-05548, a NSF Young Investigator Award DMS-9457859 and a Sloan Research Fellowship.

INTRODUCTION

Index theory and K-Theory have been close subjects since their appearance [1, 4]. Several recent index theorems that have found applications to Novikov's Conjecture use algebraic K-Theory in an essential way, as a natural target for the generalized indices that they compute. Some of these generalized indices are "von Neumann dimensions"–like in the L^2–index theorem for coverings [3] that, roughly speaking, computes the trace of the projection on the space of solutions of an elliptic differential operator on a covering space. The von Neumann dimension of the index does not fully recover the information contained in the abstract (i.e., algebraic K-Theory index) but this situation is remedied by considering "higher traces," as in the Connes–Moscovici Index Theorem for coverings [11]. (Since the appearance of this theorem, index theorems that compute the pairing between higher traces and the K–Theory class of the index are called "higher index theorems.")

In [30], a general higher index morphism (i.e., a bivariant character) was defined for a class of algebras–or, more precisely, for a class of extensions of algebras–that is large enough to accommodate most applications. However, the index theorem proved there was obtained only under some fairly restrictive conditions, too restrictive for most applications. In this paper we completely remove these restrictions using a recent breakthrough result of Cuntz and Quillen.

In [16], Cuntz and Quillen have shown that periodic cyclic homology, denoted HP_*, satisfies excision, and hence that any two–sided ideal I of a complex algebra A gives rise to a periodic six-term exact sequence

$$
\begin{tikzcd}
\mathrm{HP}_0(I) \arrow[r] & \mathrm{HP}_0(A) \arrow[r] & \mathrm{HP}_0(A/I) \arrow[d, "\partial"] \\
\mathrm{HP}_1(A/I) \arrow[u, "\partial"] & \mathrm{HP}_1(A) \arrow[l] & \mathrm{HP}_1(I) \arrow[l]
\end{tikzcd}
\tag{1}
$$

similar to the topological K–Theory exact sequence [1]. Their result generalizes earlier results from [38]. (See also [14, 15].)

If M is a smooth manifold and $A = C^\infty(M)$, then $\mathrm{HP}_*(A)$ is isomorphic to the de Rham cohomology of M, and the Chern–Connes character on (algebraic) K–Theory generalizes the Chern–Weil construction of characteristic classes using connection and curvature [10]. In view of this result, the excision property, equation (1), gives more evidence that periodic cyclic homology is the "right" extension of de Rham homology from smooth manifolds to algebras. Indeed, if $I \subset A$ is the ideal of functions vanishing on a closed submanifold $N \subset M$, then

$$\mathrm{HP}_*(I) = \mathrm{H}^*_{DR}(M, N)$$

and the exact sequence for continuous periodic cyclic homology coincides with the exact sequence for de Rham cohomology. This result extends to (not necessarily smooth) complex affine algebraic varieties [22].

The central result of this paper, Theorem 1.6, Section 1, states that the Chern–Connes character

$$ch : \mathrm{K}_i^{\mathrm{alg}}(A) \to \mathrm{HP}_i(A),$$

where $i = 0, 1$, is a natural transformation from the six term exact sequence in (lower) algebraic K–Theory to the periodic cyclic homology exact sequence. In this formulation, Theorem 1.6 generalizes the corresponding result for the Chern character on the K–Theory of compact topological spaces, thus extending the list of common features of de Rham and cyclic cohomology.

The new ingredient in Theorem 1.6, besides the naturality of the Chern–Connes character, is the compatibility between the connecting (or index) morphism in algebraic K–Theory and the boundary map in the Cuntz–Quillen exact sequence (Theorem 1.5). Because the connecting morphism

$$\mathrm{Ind} : \mathrm{K}_1^{\mathrm{alg}}(A/I) \to \mathrm{K}_0^{\mathrm{alg}}(I)$$

associated to a two-sided ideal $I \subset A$ generalizes the index of Fredholm operators, Theorem 1.5 can be regarded as an abstract "higher index theorem," and the computation of the boundary map in the periodic cyclic cohomology exact sequence can be regarded as a "cohomological index formula."

We now describe the contents of the paper in more detail.

If τ is a trace on the two–sided ideal $I \subset A$, then τ induces a morphism

$$\tau_* : \mathrm{K}_0^{\mathrm{alg}}(I) \to \mathbb{C}.$$

More generally, one can–and has to–allow τ to be a "higher trace," while still getting a morphism $\tau_* : \mathrm{K}_1^{\mathrm{alg}}(I) \to \mathbb{C}$. Our main goal in Section 1 is to identify, as explicitly as possible, the composition $\tau_* \circ \mathrm{Ind} : \mathrm{K}_1^{\mathrm{alg}}(A/I) \to \mathbb{C}$. For traces this is done in Lemma 1.1, which generalizes a formula of Fedosov. In general,

$$\tau_* \circ \mathrm{Ind} = (\partial \tau)_*,$$

where $\partial : \mathrm{HP}^0(I) \to \mathrm{HP}^1(A/I)$ is the boundary map in periodic cyclic cohomology. Since ∂ is defined purely algebraically, it is usually easier to compute it than it is to compute Ind, not to mention that the group $\mathrm{K}_0^{\mathrm{alg}}(I)$ is not known in many interesting situations, which complicates the computation of Ind even further.

In Section 2 we study the properties of ∂ and show that ∂ is compatible with various product type operations on cyclic cohomology. The proofs use cyclic vector spaces [9] and the external product $\times$ studied in [30], which generalizes the cross-product in singular homology. The most important property of ∂ is with respect to the tensor product of an exact sequence of algebras by another algebra (Theorem 2.6). We also show that the boundary map ∂ coincides with the morphism induced by the odd bivariant character constructed in [30], whenever the later is defined (Theorem 2.10).

As an application, in Section 3 we give a new proof of the Connes–Moscovici index theorem for coverings [11]. The original proof uses estimates with heat kernels. Our proof uses the results of the first two sections to reduce the Connes–Moscovici index theorem to the Atiyah–Singer index theorem for elliptic operators on compact manifolds.

The main results of this paper were announced in [32], and a preliminary version of this paper has been circulated as "Penn State preprint" no. PM 171, March 1994. Although this is a completely revised version of that preprint, the proofs have not been changed in any essential way. However, a few related preprints and papers have appeared since this paper was first written; they include [12, 13, 33].

I would like to thank Joachim Cuntz for sending me the preprints that have lead to this work and for several useful discussions. Also, I would like to thank the

Mathematical Institute of Heidelberg University for hospitality while parts of this manuscript were prepared, and to the referee for many useful comments.

1. Index theorems and Algebraic K–Theory

We begin this section by reviewing the definitions of the groups $\mathrm{K}_0^{\mathrm{alg}}$ and $\mathrm{K}_1^{\mathrm{alg}}$ and of the index morphism $\mathrm{Ind} : \mathrm{K}_1^{\mathrm{alg}}(A/I) \to \mathrm{K}_0^{\mathrm{alg}}(I)$ associated to a two-sided ideal $I \subset A$. There are easy formulas that relate these groups to Hochschild homology, and we review those as well. Then we prove an intermediate result that generalizes a formula of Fedosov in our Hochschild homology setting, which will serve both as a lemma in the proof of Theorem 1.5, and as a motivation for some of the formalisms developed in this paper. The main result of this section is the compatibility between the connecting (or index) morphism in algebraic K–Theory and the boundary morphism in cyclic cohomology (Theorem 1.5). An equivalent form of Theorem 1.5 states that the Chern–Connes character is a natural transformation from the six term exact sequence in algebraic K–Theory to periodic cyclic homology. These results extend the results in [30] in view of Theorem 2.10.

All algebras considered in this paper are complex algebras.

1.1. Pairings with traces and a Fedosov type formula. It will be convenient to define the group $\mathrm{K}_0^{\mathrm{alg}}(A)$ in terms of idempotents $e \in M_\infty(A)$, that is, in terms of matrices e satisfying $e^2 = e$. Two idempotents, e and f, are called *equivalent* (in writing, $e \sim f$) if there exist x, y such that $e = xy$ and $f = yx$. The direct sum of two idempotents, e and f, is the matrix $e \oplus f$ (with e in the upper–left corner and f in the lower–right corner). With the direct–sum operation, the set of equivalence classes of idempotents in $M_\infty(A)$ becomes a monoid denoted $\mathcal{P}(A)$. The group $\mathrm{K}_0^{\mathrm{alg}}(A)$ is defined to be the Grothendieck group associated to the monoid $\mathcal{P}(A)$. If $e \in M_\infty(A)$ is an idempotent, then the class of e in the group $\mathrm{K}_0^{\mathrm{alg}}(A)$ will be denoted $[e]$.

Let $\tau : A \to \mathbb{C}$ be a trace. We extend τ to a trace $M_\infty(A) \to \mathbb{C}$, still denoted τ, by the formula $\tau([a_{ij}]) = \sum_i \tau(a_{ii})$. If $e \sim f$, then $e = xy$ and $f = yx$ for some x and y, and then the tracial property of τ implies that $\tau(e) = \tau(f)$. Moreover $\tau(e \oplus f) = \tau(e) + \tau(f)$, and hence τ defines an additive map $\mathcal{P}(A) \to \mathbb{C}$. From the universal property of the Grothendieck group associated to a monoid, it follows that we obtain a well defined group morphism (or pairing with τ)

$$\mathrm{K}_0^{\mathrm{alg}}(A) \ni [e] \longrightarrow \tau_*([e]) = \tau(e) \in \mathbb{C}. \tag{2}$$

The pairing (2) generalizes to not necessarily unital algebras I and traces $\tau : I \to \mathbb{C}$ as follows. First, we extend τ to $I^+ = I + \mathbb{C}1$, the algebra with adjoint unit, to be zero on 1. Then, we obtain, as above, a morphism $\tau_* : \mathrm{K}_0^{\mathrm{alg}}(I^+) \to \mathbb{C}$. The morphism $\tau_* : \mathrm{K}_0^{\mathrm{alg}}(I) \to \mathbb{C}$ is obtained by restricting from $\mathrm{K}_0^{\mathrm{alg}}(I^+)$ to $\mathrm{K}_0^{\mathrm{alg}}(I)$, defined to be the kernel of $\mathrm{K}_0^{\mathrm{alg}}(I^+) \to \mathrm{K}_0^{\mathrm{alg}}(\mathbb{C})$.

The definition of $\mathrm{K}_1^{\mathrm{alg}}(A)$ is shorter:

$$\mathrm{K}_1^{\mathrm{alg}}(A) = \lim_{\to} GL_n(A)/[GL_n(A), GL_n(A)].$$

In words, $\mathrm{K}_1^{\mathrm{alg}}(A)$ is the abelianization of the group of invertible matrices of the form $1 + a$, where $a \in M_\infty(A)$. The pairing with traces is replaced by a pairing with Hochschild 1–cocycles as follows.

If $\phi : A \otimes A$ is a Hochschild 1-cocycle, then the the functional ϕ defines a morphism $\phi_* : \mathrm{K}_1^{\mathrm{alg}}(A) \to \mathbb{C}$, by first extending ϕ to matrices over A, and then by pairing it with the Hochschild 1–cycle $u \otimes u^{-1}$. Explicitly, if $u = [a_{ij}]$, with inverse $u^{-1} = [b_{ij}]$, then the morphism ϕ_* is

$$\mathrm{K}_1^{\mathrm{alg}}(A) \ni [u] \longrightarrow \phi_*([u]) = \sum_{i,j} \phi(a_{ij}, b_{ji}) \in \mathbb{C}. \tag{3}$$

The morphism ϕ_* depends only on the class of ϕ in the Hochschild homology group $\mathrm{HH}_1(A)$ of A.

If $0 \to I \to A \to A/I \to 0$ is an exact sequence of algebras, that is, if I is a two–sided ideal of A, then there exists an exact sequence [26],

$$\mathrm{K}_1^{\mathrm{alg}}(I) \to \mathrm{K}_1^{\mathrm{alg}}(A) \to \mathrm{K}_1^{\mathrm{alg}}(A/I) \xrightarrow{\mathrm{Ind}} \mathrm{K}_0^{\mathrm{alg}}(I) \to \mathrm{K}_0^{\mathrm{alg}}(A) \to \mathrm{K}_0^{\mathrm{alg}}(A/I),$$

of Abelian groups, called the *algebraic K–theory exact sequence*. The connecting (or *index*) morphism

$$\mathrm{Ind} : K_1^{alg}(A/I) \to K_0^{alg}(I)$$

will play an important role in this paper and is defined as follows. Let u be an invertible element in some matrix algebra of A/I. By replacing A/I with $M_n(A/I)$, for some large n, we may assume that $u \in A/I$. Choose an invertible element $v \in M_2(A)$ that projects to $u \oplus u^{-1}$ in $M_2(A/I)$, and let $e_0 = 1 \oplus 0$ and $e_1 = v e_0 v^{-1}$. Because $e_1 \in M_2(I^+)$, the idempotent e_1 defines a class in $\mathrm{K}_0^{\mathrm{alg}}(I^+)$. Since $e_1 - e_0 \in M_2(I)$, the difference $[e_1] - [e_0]$ is actually in $\mathrm{K}_0^{\mathrm{alg}}(I)$ and depends only on the class $[u]$ of u in $\mathrm{K}_1^{\mathrm{alg}}(A/I)$. Finally, we define

$$\mathrm{Ind}([u]) = [e_1] - [e_0]. \tag{4}$$

To obtain an explicit formula for e_1, choose liftings $a, b \in A$ of u and u^{-1} and let v, the lifting, to be the matrix

$$v = \begin{bmatrix} 2a - aba & ab - 1 \\ 1 - ba & b \end{bmatrix},$$

as in [26], page 22. Then a short computation gives

$$e_1 = \begin{bmatrix} 2ab - (ab)^2 & a(2 - ba)(1 - ba) \\ (1 - ba)b & (1 - ba)^2 \end{bmatrix}. \tag{5}$$

Continuing the study of the exact sequence $0 \to I \to A \to A/I \to 0$, choose an arbitrary linear lifting, $l : A/I^2 \to A$. If τ is a trace on I, we let

$$\phi_\tau(a, b) = \tau([l(a), l(b)] - l([a, b])). \tag{6}$$

Because $[a, xy] = [ax, y] + [ya, x]$, we have $\tau([A, I^2]) = 0$, and hence ϕ_τ is a Hochschild 1–cocycle on A/I^2 (i.e., $\phi_\tau(ab, c) - \phi_\tau(a, bc) + \phi_\tau(ca, b)$). The class of ϕ_τ in $\mathrm{HH}^1(A/I^2)$, denoted $\partial\tau$, turns out to be independent of the lifting l. If A is a locally convex algebra, then we assume that we can choose the lifting l to be continuous. If $\tau([A, I]) = 0$, then it is enough to consider a lifting of $A \to A/I$.

The morphisms $(\partial\tau)_* : \mathrm{K}_1^{\mathrm{alg}}(A/I^2) \to \mathbb{C}$ and $\tau_* : \mathrm{K}_0^{\mathrm{alg}}(I^2) \to \mathbb{C}$ are related through the following lemma.

LEMMA. 1.1. *Let τ be a trace on a two-sided ideal $I \subset A$. If*

$$\mathrm{Ind} : \mathrm{K}_1^{\mathrm{alg}}(A/I^2) \to \mathrm{K}_0^{\mathrm{alg}}(I^2)$$

is the connecting morphism of the algebraic K–Theory exact sequence associated to the two-sided ideal I^2 of A, then

$$\tau_* \circ \mathrm{Ind} = (\partial\tau)_*.$$

If $\tau([A, I]) = 0$, then we may replace I^2 by I.

Proof. We check that $\tau_* \circ \mathrm{Ind}([u]) = (\partial\tau)_*([u])$, for each invertible $u \in M_n(A/I^2)$. By replacing A/I^2 with $M_n(A/I^2)$, we may assume that $n = 1$.

Let $l : A/I^2 \to A$ be the linear lifting used to define the 1–cocycle ϕ_τ representing $\partial\tau$, equation (6), and choose $a = l(u)$ and $b = l(u^{-1})$ in the formula for e_1, equation (5). Then, the left hand side of our formula becomes

$$\tau_*\big(\mathrm{Ind}([u])\big) = \tau\big((1-ba)^2\big) - \tau\big((1-ab)^2\big) = 2\tau([a,b]) - \tau([a, bab]). \tag{7}$$

Because $(1-ba)b$ is in I^2, we have $\tau([a, bab]) = \tau([a, b])$, and hence

$$\tau_*(\mathrm{Ind}([u])) = \tau_*([e_1] - [e_0]) = \tau(e_1 - e_0) = \tau([a, b]).$$

Since the right hand side of our formula is

$$(\partial\tau)_*([u]) = (\partial\tau)(u, u^{-1}) = \tau([l(u), l(u^{-1})] - l([u, u^{-1}])) = \tau([a, b]),$$

the proof is complete. □

Lemma 1.1 generalizes a formula of Fedosov in the following situation. Let $\mathcal{B}(\mathcal{H})$ be the algebra of bounded operators on a fixed separable Hilbert space $\mathcal{H}$ and $\mathcal{C}_p(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ be the (non-closed) ideal of p–summable operators [36] on $\mathcal{H}$:

$$\mathcal{C}_p(\mathcal{H}) = \{A \in \mathcal{B}(\mathcal{H}),\ Tr(A^*A)^{p/2} < \infty\}. \tag{8}$$

(We will sometimes omit $\mathcal{H}$ and write simply $\mathcal{C}_p$ instead of $\mathcal{C}_p(\mathcal{H})$.) Suppose now that the algebra A consists of bounded operators, that $I \subset \mathcal{C}_1$, and that a is an element of A whose projection u in A/I is invertible. Then a is a *Fredholm* operator, and, for a suitable choice of a lifting b of u^{-1}, the operators $1 - ba$ and $1 - ab$ become the orthogonal projection onto the kernel of a and, respectively, the kernel of a^*. Finally, if $\tau = Tr$, this shows that

$$Tr_*\big(\mathrm{Ind}([u])\big) = \dim\ker(a) - \dim\ker(a^*)$$

and hence that $Tr_* \circ \mathrm{Ind}$ recovers the Fredholm index of a. (The *Fredholm index* of a, denoted $ind(a)$, is by definition the right-hand side of the above formula.) By equation (7), we see that we also recover a form of Fedosov's formula:

$$ind(a) = Tr\big((1-ba)^k\big) - Tr\big((1-ab)^k\big)$$

if b is an inverse of a modulo $\mathcal{C}_p(\mathcal{H})$ and $k \geq p$.

The connecting (or boundary) morphism in the algebraic K–Theory exact sequence is usually denoted by '∂'. However, in the present paper, this notation becomes unsuitable because the notation '∂' is reserved for the boundary morphism in the periodic cyclic cohomology exact sequence. Besides, the notation 'Ind' is supposed to suggest the name 'index morphism' for the connecting morphism in the algebraic K–Theory exact sequence, a name justified by the relation that exists between Ind and the indices of Fredholm operators, as explained above.

1.2. "Higher traces" and excision in cyclic cohomology. The example of $A = C^\infty(M)$, for M a compact smooth manifold, shows that, in general, few morphisms $\mathrm{K}_0^{\mathrm{alg}}(A) \to \mathbb{C}$ are given by pairings with traces. This situation is corrected by considering 'higher-traces,' [10].

Let A be a unital algebra and

$$b'(a_0 \otimes \ldots \otimes a_n) = \sum_{i=0}^{n-1} (-1)^i a_0 \otimes \ldots \otimes a_i a_{i+1} \otimes \ldots \otimes a_n,$$
$$b(a_0 \otimes \ldots \otimes a_n) = b'(a_0 \otimes \ldots \otimes a_n) + (-1)^n a_n a_0 \otimes \ldots \otimes a_{n-1}, \tag{9}$$

for $a_i \in A$. The *Hochschild homology* groups of A, denoted $\mathrm{HH}_*(A)$, are the homology groups of the complex $(A \otimes (A/\mathbb{C}1)^{\otimes n}, b)$. The *cyclic homology* groups [10, 24, 37] of a unital algebra A, denoted $\mathrm{HC}_n(A)$, are the homology groups of the complex $(\mathcal{C}(A), b+B)$, where

$$\mathcal{C}_n(A) = \bigoplus_{k \geq 0} A \otimes (A/\mathbb{C}1)^{\otimes n-2k}. \tag{10}$$

b is the Hochschild homology boundary map, equation (9), and B is defined by

$$B(a_0 \otimes \ldots \otimes a_n) = s \sum_{k=0}^{n} t^k (a_0 \otimes \ldots \otimes a_n). \tag{11}$$

Here we have used the notation of [10], that $s(a_0 \otimes \ldots \otimes a_n) = 1 \otimes a_0 \otimes \ldots \otimes a_n$ and $t(a_0 \otimes \ldots \otimes a_n) = (-1)^n a_n \otimes a_0 \otimes \ldots \otimes a_{n-1}$.

More generally, Hochschild and cyclic homology groups can be defined for "mixed complexes," [21]. A *mixed complex* $(\mathcal{X}, b, B)$ is a graded vector space $(\mathcal{X}_n)_{n \geq 0}$, endowed with two differentials b and B, $b : \mathcal{X}_n \to \mathcal{X}_{n-1}$ and $B : \mathcal{X}_n \to \mathcal{X}_{n+1}$, satisfying the compatibility relation $b^2 = B^2 = bB + Bb = 0$. The *cyclic complex*, denoted $\mathcal{C}(\mathcal{X})$, associated to a mixed complex $(\mathcal{X}, b, B)$ is the complex

$$\mathcal{C}_n(\mathcal{X}) = \mathcal{X}_n \oplus \mathcal{X}_{n-2} \oplus \mathcal{X}_{n-4} \ldots = \bigoplus_{k \geq 0} \mathcal{X}_{n-2k},$$

with differential $b + B$. The cyclic homology groups of the mixed complex $\mathcal{X}$ are the homology groups of the cyclic complex of $\mathcal{X}$:

$$\mathrm{HC}_n(\mathcal{X}) = \mathrm{H}_n(\mathcal{C}(\mathcal{X}), b+B).$$

Cyclic cohomology is defined to be the homology of the complex

$$(\mathcal{C}(\mathcal{X})' = \mathrm{Hom}(\mathcal{C}(\mathcal{X}), \mathbb{C}), (b+B)'),$$

dual to $\mathcal{C}(\mathcal{X})$. From the form of the cyclic complex it is clear that there exists a morphism $S : \mathcal{C}_n(\mathcal{X}) \to \mathcal{C}_{n-2}(\mathcal{X})$. We let

$$\mathcal{C}_n(\mathcal{X}) = \lim_{\leftarrow} \mathcal{C}_{n+2k}(\mathcal{X})$$

as $k \to \infty$, the inverse system being with respect to the periodicity operator $\mathbb{S}$. Then the *periodic cyclic homology* of $\mathcal{X}$ (respectively, the *periodic cyclic cohomology* of $\mathcal{X}$), denoted $\mathrm{HP}_*(\mathcal{X})$ (respectively, $\mathrm{HP}^*(\mathcal{X})$) is the homology (respectively, the cohomology) of $\mathcal{C}_n(\mathcal{X})$ (respectively, of the complex $\lim_{\rightarrow} \mathcal{C}_{n+2k}(\mathcal{X})'$).

If A is a unital algebra, we denote by $\mathcal{X}(A)$ the mixed complex obtained by letting $\mathcal{X}_n(A) = A \otimes (A/\mathbb{C}1)^{\otimes n}$ with differentials b and B given by (9) and (11). The various homologies of $\mathcal{X}(A)$ will not include $\mathcal{X}$ as part of notation. For example, the periodic cyclic homology of $\mathcal{X}$ is denoted $\mathrm{HP}_*(A)$.

For a topological algebra A we may also consider *continuous* versions of the above homologies by replacing the ordinary tensor product with the projective tensor product. We shall be especially interested in the continuous cyclic cohomology of A, denoted $\mathrm{HP}^*_{\mathrm{cont}}(A)$. An important example is $A = C^\infty(M)$, for a compact smooth manifold M. Then the Hochschild-Kostant-Rosenberg map

$$\chi : A^{\hat{\otimes} n+1} \ni a_0 \otimes a_1 \otimes \ldots \otimes a_n \longrightarrow (n!)^{-1} a_0 da_1 \ldots da_n \in \Omega^n(M) \tag{12}$$

to smooth forms gives an isomorphism

$$\mathrm{HP}_i^{\mathrm{cont}}(C^\infty(M)) \simeq \bigoplus_k \mathrm{H}_{DR}^{i+2k}(M)$$

of *continuous* periodic cyclic homology with the de Rham cohomology of M [10, 24] made $\mathbb{Z}_2$–periodic. The normalization factor $(n!)^{-1}$ is convenient because it transforms B into the de Rham differential d_{DR}. It is also the right normalization as far as Chern characters are involved, and it is also compatible with products, Theorem 3.5. From now on, we shall use the de Rham's Theorem

$$\mathrm{H}^i_{DR}(M) \simeq \mathrm{H}^i(M)$$

to identify de Rham cohomology and singular cohomology *with complex coefficients* of the compact manifold M.

Sometimes we will use a version of continuous periodic cyclic cohomology for algebras A that have a locally convex space structure, but for which the multiplication is only partially continuous. In that case, however, the tensor products $A^{\otimes n+1}$ come with natural topologies, for which the differentials b and B are continuous. This is the case for some of the groupoid algebras considered in the last section. The periodic cyclic cohomology is then defined using continuous multi-linear cochains.

One of the original descriptions of cyclic cohomology was in terms of "higher traces" [10]. A *higher trace*–or cyclic cocycle–is a continuous multilinear map $\phi : A^{\otimes n+1} \to \mathbb{C}$ satisfying $\phi \circ b = 0$ and $\phi(a_1, \ldots, a_n, a_0) = (-1)^n \phi(a_0, \ldots, a_n)$. Thus cyclic cocycles are, in particular, Hochschild cocycles. The last property, the cyclic invariance, justifies the name "cyclic cocycles." The other name, "higher traces" is justified since cyclic cocycles on A define traces on the universal differential graded algebra of A.

If $I \subset A$ is a two–sided ideal, we denote by $\mathcal{C}(A, I)$ the kernel of $\mathcal{C}(A) \to \mathcal{C}(A/I)$. For possibly non-unital algebras I, we define the cyclic homology of I using the complex $\mathcal{C}(I^+, I)$. The cyclic cohomology and the periodic versions of these groups are defined analogously, using $\mathcal{C}(I^+, I)$. For topological algebras we replace the algebraic tensor product by the projective tensor product.

An equivalent form of the excision theorem in periodic cyclic cohomology is the following result.

THEOREM. 1.2 (Cuntz–Quillen). *The inclusion $\mathcal{C}(I^+, I) \hookrightarrow \mathcal{C}(A, I)$ induces an isomorphism, $\mathrm{HP}^*(A, I) \simeq \mathrm{HP}^*(I)$, of periodic cyclic cohomology groups.*

This theorem is implicit in [16], and follows directly from the proof there of the Excision Theorem by a sequence of commutative diagrams, using the Five Lemma each time.[2]

This alternative definition of excision sometimes leads to explicit formulae for ∂. We begin by observing that the short exact sequence of complexes $0 \to \mathcal{C}(A, I) \to \mathcal{C}(A) \to \mathcal{C}(A/I) \to 0$ defines a long exact sequence

$$.. \leftarrow \mathrm{HP}^n(A, I) \leftarrow \mathrm{HP}^n(A) \leftarrow \mathrm{HP}^n(A/I) \xleftarrow{\partial} \mathrm{HP}^{n-1}(A, I) \leftarrow \mathrm{HP}^{n-1}(A) \leftarrow ..$$

in cyclic cohomology that maps naturally to the long exact sequence in periodic cyclic cohomology.

Most important for us, the boundary map $\partial : \mathrm{HP}^n(A, I) \to \mathrm{HP}^{n+1}(A/I)$ is determined by a standard algebraic construction. We now want to prove that this boundary morphism recovers a previous construction, equation (6), in the particular case $n = 0$. As we have already observed, a trace $\tau : I \to \mathbb{C}$ satisfies $\tau([A, I^2]) = 0$, and hence defines by restriction an element of $\mathrm{HC}^0(A, I^2)$. The traces are the cycles of the group $\mathrm{HC}^0(I)$, and thus we obtain a linear map $\mathrm{HC}^0(I) \to \mathrm{HC}^0(A, I^2)$. From the definition of $\partial : \mathrm{HP}^0(A, I) \to \mathrm{HP}^1(A/I)$, it follows that $\partial[\tau]$ is the class of the cocycle $\phi(a, b) = \tau([l(a), l(b)] - l([a, b]))$, which is cyclically invariant, by construction. (Since our previous notation for the class of ϕ was $\partial\tau$, we have thus obtained the paradoxical relation $\partial[\tau] = \partial\tau$; we hope this will not cause any confusions.)

Below we shall also use the natural map (transformation)

$$\mathrm{HC}^n \to \mathrm{HP}^n = \lim_{k \to \infty} \mathrm{HC}^{n+2k}.$$

LEMMA. 1.3. *The diagram*

$$\begin{tikzcd}
\mathrm{HC}^0(I) \arrow[r] \arrow[d] & \mathrm{HC}^0(A, I^2) \arrow[r, "\partial"] \arrow[d] & \mathrm{HC}^1(A/I^2) \arrow[d] & \mathrm{HC}^1(A/I) \arrow[l] \arrow[d] \\
\mathrm{HP}^0(I) \arrow[r, "\sim"] & \mathrm{HP}^0(A, I^2) \arrow[r, "\partial"] & \mathrm{HP}^1(A/I^2) & \mathrm{HP}^1(A/I) \arrow[l, "\sim"']
\end{tikzcd}$$

commutes. Consequently, if $\tau \in \mathrm{HC}^0(I)$ is a trace on I and $[\tau] \in \mathrm{HP}^0(I)$ is its class in periodic cyclic homology, then $\partial[\tau] = [\partial\tau] \in \mathrm{HP}^1(A/I)$, where $\partial\tau \in \mathrm{HC}^1(A/I^2)$ is given by the class of the cocycle ϕ defined in equation (6) *(see also above).*

Proof. The commutativity of the diagram follows from definitions. If we start with a trace $\tau \in \mathrm{HC}^0(I)$ and follow counterclockwise through the diagram from the upper–left corner to the lower–right corner we obtain $\partial[\tau]$; if we follow clockwise, we obtain the description for $\partial[\tau]$ indicated in the statement. ☐

1.3. AN ABSTRACT "HIGHER INDEX THEOREM". We now generalize Lemma 1.1 to periodic cyclic cohomology. Recall that the pairings (2) and (3) have been generalized to pairings

$$\mathrm{K}_i^{\mathrm{alg}}(A) \otimes \mathrm{HC}^{2n+i}(A) \longrightarrow \mathbb{C}, \quad i = 0, 1.$$

[10]. Thus, if ϕ be a higher trace representing a class $[\phi] \in \mathrm{HC}^{2n+i}(A)$, then, using the above pairing, ϕ defines morphisms $\phi_* : \mathrm{K}_i^{\mathrm{alg}}(A) \to \mathbb{C}$, where $i = 0, 1$. The explicit formulae for these morphisms are $\phi_*([e]) = (-1)^n \frac{(2n)!}{n!}\phi(e, e, \ldots, e)$, if $i = 0$ and e

[2] I am indebted to Joachim Cuntz for pointing out this fact to me.

is an idempotent, and $\phi_*([u]) = (-1)^n n!\phi_*(u, u^{-1}, u, \ldots, u^{-1})$, if $i = 1$ and u is an invertible element. The constants in these pairings are meaningful and are chosen so that these pairings are compatible with the periodicity operator.

Consider the standard orthonormal basis $(e_n)_{n\geq 0}$ of the space $l^2(\mathbb{N})$ of square summable sequences of complex numbers; the *shift* operator S is defined by $Se_n = e_{n+1}$. The adjoint S^* of S then acts by $S^*e_0 = 0$ and $S^*e_{n+1} = e_n$, for $n \geq 0$. The operators S and S^* are related by $S^*S = 1$ and $SS^* = 1-p$, where p is the orthogonal projection onto the vector space generated by e_0.

Let $\mathcal{T}$ be the algebra generated by S and S^* and $\mathbb{C}[w, w^{-1}]$ be the algebra of Laurent series in the variable w, $\mathbb{C}[w, w^{-1}] = \{\sum_{n=-N}^{N} a_k w^k,\ a_k \in \mathbb{C}\} \simeq \mathbb{C}[\mathbb{Z}]$. Then there exists an exact sequence

$$0 \to M_\infty(\mathbb{C}) \to \mathcal{T} \to \mathbb{C}[w, w^{-1}] \to 0,$$

called *the Toeplitz extension*, which sends S to w and S^* to w^{-1}.

Let $\mathbb{C}\langle a, b\rangle$ be the free *non-commutative* unital algebra generated by the symbols a and b and $J = \ker(\mathbb{C}\langle a, b\rangle \to \mathbb{C}[w, w^{-1}])$, the kernel of the unital morphism that sends $a \to w$ and $b \to w^{-1}$. Then there exists a morphism $\psi_0 : \mathbb{C}\langle a, b\rangle \to \mathcal{T}$, uniquely determined by $\psi_0(a) = S$ and $\psi_0(b) = S^*$, which defines, by restriction, a morphism $\psi : J \to M_\infty(\mathbb{C})$, and hence a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & J \arrow[r] \arrow[d, "\psi"'] & \mathbb{C}\langle a, b\rangle \arrow[r] \arrow[d, "\psi_0"] & \mathbb{C}[w, w^{-1}] \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & M_\infty(\mathbb{C}) \arrow[r] & \mathcal{T} \arrow[r] & \mathbb{C}[w, w^{-1}] \arrow[r] & 0
\end{tikzcd}$$

LEMMA. 1.4. *Using the above notations, we have that* $\mathrm{HC}^*(J)$ *is singly generated by the trace* $\tau = Tr \circ \psi$.

Proof. We know that $\mathrm{HP}^i(\mathbb{C}[w, w^{-1}]) \simeq \mathbb{C}$, see [24]. Then Lemma 1.1, Lemma 1.3, and the exact sequence in periodic cyclic cohomology prove the vanishing of the *reduced* periodic cyclic cohomology groups:

$$\widetilde{\mathrm{HC}}^*(\mathcal{T}) = \ker(\mathrm{HP}^*(\mathcal{T}) \to \mathrm{HP}^*(\mathbb{C})).$$

The algebra $\mathbb{C}\langle a, b\rangle$ is the tensor algebra of the vector space $\mathbb{C}a \oplus \mathbb{C}b$, and hence the groups $\widetilde{\mathrm{HC}}^*(T(V))$ also vanish [24]. It follows that the morphism ψ_0 induces (trivially) an isomorphism in cyclic cohomology. The comparison morphism between the Cuntz–Quillen exact sequences associated to the two extensions shows, using "the Five Lemma," that the induced morphisms $\psi^* : \mathrm{HP}^*(M_\infty(\mathbb{C})) \to \mathrm{HP}^*(J)$ is also an isomorphism. This proves the result since the canonical trace Tr generates $\mathrm{HP}^*(M_\infty(\mathbb{C}))$. □

We are now ready to state the main result of this section, the compatibility of the boundary map in the periodic cyclic cohomology exact sequence with the index (i.e., connecting) map in the algebraic K–Theory exact sequence. The following theorem generalizes Theorem 5.4 from [30].

THEOREM. 1.5. *Let* $0 \to I \to A \to A/I \to 0$ *be an exact sequence of complex algebras, and let* $\mathrm{Ind} : \mathrm{K}_1^{\mathrm{alg}}(A/I) \to \mathrm{K}_0^{\mathrm{alg}}(I)$ *and* $\partial : \mathrm{HP}^0(I) \to \mathrm{HP}^1(A/I)$ *be the* *connecting morphisms in algebraic K–Theory and, respectively, in periodic cyclic cohomology. Then, for any $\varphi \in \mathrm{HP}^0(I)$ and $[u] \in \mathrm{K}_1^{\mathrm{alg}}(A/I)$, we have*

$$\varphi_*(\mathrm{Ind}[u]) = (\partial\varphi)_*[u]\,. \tag{13}$$

Proof. We begin by observing that if the class of φ can be represented by a trace (that is, if φ is the equivalence class of a trace in the group $\mathrm{HP}^0(I)$) then the boundary map in periodic cyclic cohomology is computed using the recipe we have indicated, Lemma 1.3, and hence the result follows from Lemma 1.1. In particular, the theorem is true for the exact sequence

$$0 \longrightarrow J \to \mathbb{C}\langle\, a, b\,\rangle \to \mathbb{C}[w, w^{-1}] \longrightarrow 0,$$

because all classes in $\mathrm{HP}^0(J)$ are defined by traces, as shown in Lemma 1.4. We will now show that this particular case is enough to prove the general case "by universality."

Let u be an invertible element in $M_n(A/I)$. After replacing the algebras involved by matrix algebras, if necessary, we may assume that $n = 1$, and hence that u is an invertible element in A/I. This invertible element then gives rise to a morphism $\eta : \mathbb{C}[w, w^{-1}] \to A/I$ that sends w to u. A choice of liftings $a_0, b_0 \in A$ of u and u^{-1} defines a morphism $\psi_0 : \mathbb{C}\langle\, a, b\,\rangle \to A$, uniquely determined by $\psi_0(a) = a_0$ and $\psi_0(b) = b_0$, which restricts to a morphism $\psi : J \to I$. In this way we obtain a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & J \arrow[r] \arrow[d, "\psi"'] & \mathbb{C}\langle\, a, b\,\rangle \arrow[r] \arrow[d, "\psi_0"] & \mathbb{C}[w, w^{-1}] \arrow[r] \arrow[d, "\eta"] & 0 \\
0 \arrow[r] & I \arrow[r] & A \arrow[r] & A/I \arrow[r] & 0
\end{tikzcd}$$

of algebras and morphisms.

We claim that the naturality of the index morphism in algebraic K–Theory and the naturality of the boundary map in periodic cyclic cohomology, when applied to the above exact sequence, prove the theorem. Indeed, we have

$$\psi_* \circ \mathrm{Ind} = \mathrm{Ind} \circ \eta_* : \mathrm{K}_1^{\mathrm{alg}}(\mathbb{C}[w, w^{-1}]) \to \mathrm{K}_0^{\mathrm{alg}}(I), \quad \text{and}$$

$$\partial \circ \psi^* = \eta^* \circ \partial : \mathrm{HP}^*(I) \to \mathrm{HP}^{*+1}(\mathbb{C}[w, w^{-1}]).$$

As observed in the beginning of the proof, the theorem is true for the cocycle $\psi^*(\varphi)$ on J, and hence $(\psi^*(\varphi))_*(\mathrm{Ind}\,[w]) = (\partial \circ \psi^*(\varphi))_*[w]$. Finally, from definition, we have that $\eta_*[w] = [u]$. Combining these relations we obtain

$$\begin{aligned}
\varphi_*(\mathrm{Ind}\,[u]) &= \varphi_*(\mathrm{Ind} \circ \eta_*[w]) = \varphi_*(\psi_* \circ \mathrm{Ind}[w]) = (\psi^*(\varphi))_*(\mathrm{Ind}\,[w]) = \\
&= (\partial \circ \psi^*(\varphi))_*[w] = (\eta^* \circ \partial(\varphi))_*[w] = (\partial\varphi)_*(\eta_*[w]) = (\partial\varphi)_*[u].
\end{aligned}$$

The proof is complete. □

The theorem we have just proved can be extended to topological algebras and topological K–Theory. If the topological algebras considered satisfy Bott periodicity, then an analogous compatibility with the other connecting morphism can be proved and one gets a natural transformation from the six-term exact sequence in topological K–Theory to the six-term exact sequence in periodic cyclic homology. However, a factor of $2\pi\imath$ has to be taken into account because the Chern-Connes character is not directly compatible with periodicity [30], but introduces a factor of $2\pi\imath$. See [12] for details.

So far all our results have been formulated in terms of cyclic <u>co</u>homology, rather than cyclic homology. This is justified by the application in Section 3 that will use this form of the results. This is not possible, however, for the following theorem, which states that the Chern character in periodic cyclic homology (i.e., the Chern–Connes character) is a natural transformation from the six term exact sequence in (lower) algebraic K–Theory to the exact sequence in cyclic homology.

THEOREM. 1.6. *The diagram*

$$\begin{tikzcd}
\mathrm{K}_1^{\mathrm{alg}}(I) \arrow[r] \arrow[d] & \mathrm{K}_1^{\mathrm{alg}}(A) \arrow[r] \arrow[d] & \mathrm{K}_1^{\mathrm{alg}}(A/I) \arrow[r, "\mathrm{Ind}"] \arrow[d] & \mathrm{K}_0^{\mathrm{alg}}(I) \arrow[r] \arrow[d] & \mathrm{K}_0^{\mathrm{alg}}(A) \arrow[r] \arrow[d] & \mathrm{K}_0^{\mathrm{alg}}(A/I) \arrow[d] \\
\mathrm{HP}_1(I) \arrow[r] & \mathrm{HP}_1(A) \arrow[r] & \mathrm{HP}_1(A/I) \arrow[r, "\partial"] & \mathrm{HP}_0(I) \arrow[r] & \mathrm{HP}_0(A) \arrow[r] & \mathrm{HP}_0(A/I),
\end{tikzcd}$$

in which the vertical arrows are induced by the Chern characters $ch : \mathrm{K}_i^{\mathrm{alg}} \to \mathrm{HP}_i$, for $i = 0, 1$, commutes.

Proof. Only the relation $ch \circ \mathrm{Ind} = \partial \circ ch$ needs to be proved, and this is dual to Theorem 1.5. $\square$

2. PRODUCTS AND THE BOUNDARY MAP IN PERIODIC CYCLIC COHOMOLOGY

Cyclic vector spaces are a generalization of simplicial vector spaces, with which they share many features, most notably, for us, a similar behavior with respect to products.

2.1. CYCLIC VECTOR SPACES. We begin this section with a review of a few needed facts about the cyclic category Λ from [9] and [30]. We will be especially interested in the $\times$–product in bivariant cyclic cohomology. More results can be found in [23].

DEFINITION. 2.1. *The cyclic category, denoted Λ, is the category whose objects are $\Lambda_n = \{0, 1, \dots, n\}$, where $n = 0, 1, \dots$ and whose morphisms $\mathrm{Hom}_\Lambda(\Lambda_n, \Lambda_m)$ are the homotopy classes of increasing, degree one, continuous functions $\varphi : S^1 \to S^1$ satisfying $\varphi(\mathbb{Z}_{n+1}) \subseteq \mathbb{Z}_{m+1}$.*

A *cyclic vector space* is a contravariant functor from Λ to the category of complex vector spaces [9]. Explicitly, a cyclic vector space X is a graded vector space, $X = (X_n)_{n \geq 0}$, with structural morphisms $d_n^i : X_n \to X_{n-1}$, $s_n^i : X_n \to X_{n+1}$, for $0 \leq i \leq n$, and $t_{n+1} : X_n \to X_n$ such that (X_n, d_n^i, s_n^i) is a simplicial vector space ([25], Chapter VIII,§5) and t_{n+1} defines an action of the cyclic group $\mathbb{Z}_{n+1}$ satisfying $d_n^0 t_{n+1} = d_n^n$ and $s_n^0 t_{n+1} = t_{n+2}^2 s_n^n$, $d_n^i t_{n+1} = t_n d_n^{i-1}$, and $s_n^i t_{n+1} = t_{n+2} s_n^{i-1}$ for $1 \leq i \leq n$. Cyclic vector spaces form a category.

The cyclic vector space associated to a unital locally convex complex algebra A is $A^\natural = (A^{\otimes n+1})_{n \geq 0}$, with the structural morphisms

$$\begin{aligned}
s_n^i(a_0 \otimes \dots \otimes a_n) &= a_0 \otimes \dots \otimes a_i \otimes 1 \otimes a_{i+1} \otimes \dots \otimes a_n, \\
d_n^i(a_0 \otimes \dots \otimes a_n) &= a_0 \otimes \dots \otimes a_i a_{i+1} \otimes \dots \otimes a_n, \qquad \text{for } 0 \leq i < n, \text{ and} \\
d_n^n(a_0 \otimes \dots \otimes a_n) &= a_n a_0 \otimes \dots \otimes a_i a_{i+1} \otimes \dots \otimes a_{n-1}, \\
t_{n+1}(a_0 \otimes \dots \otimes a_n) &= a_n \otimes a_0 \otimes a_1 \otimes \dots \otimes a_{n-1}.
\end{aligned}$$

If $X = (X_n)_{n\geq 0}$ and $Y = (Y_n)_{n\geq 0}$ are cyclic vector spaces, then we can define on $(X_n \otimes Y_n)_{n\geq 0}$ the structure of a cyclic space with structural morphisms given by the diagonal action of the corresponding structural morphisms, s_n^i, d_n^i, and t_{n+1}, of X and Y. The resulting cyclic vector space will be denoted $X \times Y$ and called the *external product* of X and Y. In particular, we obtain that $(A\otimes B)^\natural = A^\natural \times B^\natural$ for all unital algebras A and B, and that $X \times \mathbb{C}^\natural \simeq X$ for all cyclic vector spaces X. There is an obvious variant of these constructions for locally convex algebras, obtained by using the complete projective tensor product.

The cyclic cohomology groups of an algebra A can be recovered as *Ext*–groups. For us, the most convenient definition of *Ext* is using exact sequences (or resolutions). Consider the set $E = (M_k)_{k=0}^n$ of resolutions of length $n+1$ of X by cyclic vector spaces, such that $M_n = Y$. Thus we consider exact sequences

$$E: \quad 0 \to Y = M_n \to M_{n-1} \to \cdots \to M_0 \to X \to 0,$$

of cyclic vector spaces. For two such resolutions, E and E', we write $E \simeq E'$ whenever there exists a morphism of complexes $E \to E'$ that induces the identity on X and Y. Then $\mathrm{Ext}^n_\Lambda(X,Y)$ is, by definition, the set of equivalence classes of resolutions $E = (M_k)_{k=0}^n$ with respect to the equivalence relation generated by $\simeq$. The set $\mathrm{Ext}^n_\Lambda(X,Y)$ has a natural group structure. The equivalence class in $\mathrm{Ext}^n_\Lambda(X,Y)$ of a resolution $E = (M_k)_{k=0}^n$ is denoted $[E]$. This definition of *Ext* coincides with the usual one–using resolutions by projective modules–because cyclic vector spaces form an Abelian category with enough projectives.

Given a cyclic vector space $X = (X_n)_{n\geq 0}$ define $b, b' : X_n \to X_{n-1}$ by $b' = \sum_{j=0}^{n-1}(-1)^j d_j$, $b = b' + (-1)^n d_n$. Let $s_{-1} = s_n^n \circ t_{n+1}$ be the 'extra degeneracy' of X, which satisfies $s_{-1}b' + b's_{-1} = 1$. Also let $\epsilon = 1-(-1)^n t_{n+1}$, $N = \sum_{j=0}^n (-1)^{nj} t_{n+1}^j$ and $B = \epsilon s_{-1} N$. Then (X, b, B) is a mixed complex and hence $\mathrm{HC}_*(X)$, *the cyclic homology of* X, is the homology of $(\oplus_{k\geq 0} X_{n-2k}, b+B)$, by definition. Cyclic cohomology is obtained by dualization, as before.

The *Ext*–groups recover the cyclic cohomology of an algebra A via a natural isomorphism,

$$\mathrm{HC}^n(A) \simeq \mathrm{Ext}^n_\Lambda(A^\natural, \mathbb{C}^\natural), \tag{14}$$

[9]. This isomorphism allows us to use the theory of derived functors to study cyclic cohomology, especially products.

The *Yoneda product*,

$$\mathrm{Ext}^n_\Lambda(X,Y) \otimes \mathrm{Ext}^m_\Lambda(Y,Z) \ni \xi \otimes \zeta \to \zeta \circ \xi \in \mathrm{Ext}^{n+m}_\Lambda(X,Z),$$

is defined by splicing [18]. If $E = (M_k)_{k=0}^n$ is a resolution of X, and $E' = (M'_k)_{k=0}^m$ a resolution of Y, such that $M_n = Y$ and $M'_m = Z$, then $E' \circ E$ is represented by

$$\begin{tikzcd}
0 \arrow[r] & Z = M'_m \arrow[r] & M'_{m-1} \arrow[r] & \cdots \arrow[r] & M'_0 \arrow[r] \arrow[d] & M_{n-1} \arrow[r] & \cdots \arrow[r] & M_0 \arrow[r] & X \arrow[r] & 0 \\
 & & & & Y \arrow[ur] & & & & &
\end{tikzcd}$$

The resulting product generalizes the composition of functions. Using the same notation, the *external product* $E \times E'$ is the resolution

$$E \times E' = \left(\sum_{k+j=l} M'_k \times M_j \right)_{l=0}^{n+m} .$$

Passing to equivalence classes, we obtain a product

$$\mathrm{Ext}_\Lambda^m(X,Y) \otimes \mathrm{Ext}_\Lambda^n(X_1,Y_1) \xrightarrow{\times} \mathrm{Ext}_\Lambda^{m+n}(X \times X_1, Y \times Y_1).$$

If $f : X \to X'$ is a morphism of cyclic vector spaces then we shall sometimes denote $E' \circ f = f^*(E')$, for $E' \in \mathrm{Ext}_\Lambda^n(X', \mathbb{C}^\natural)$.

The Yoneda product, "$\circ$," and the external product, "$\times$," are both associative and are related by the following identities, [30], Lemma 1.2.

LEMMA. 2.2. *Let $x \in \mathrm{Ext}_\Lambda^n(X,Y)$, $y \in \mathrm{Ext}_\Lambda^m(X_1,Y_1)$, and τ be the natural transformation $\mathrm{Ext}_\Lambda^{m+n}(X_1 \times X, Y_1 \times Y) \to \mathrm{Ext}_\Lambda^{m+n}(X \times X_1, Y \times Y_1)$ that interchanges the factors. Then*

$$x \times y = (id_Y \times y) \circ (x \times id_{X_1}) = (-1)^{mn} \ (x \times id_{Y_1}) \circ (id_X \times y),$$

$$id_X \times (y \circ z) = (id_X \times y) \circ (id_X \times z),$$

$$x \times y = (-1)^{mn} \tau(y \times x), \qquad \textit{and} \qquad x \times id_{\mathbb{C}^\natural} = x = id_{\mathbb{C}^\natural} \times x.$$

We now turn to the definition of the periodicity operator. A choice of a generator σ of the group $\mathrm{Ext}_\Lambda^2(\mathbb{C}^\natural, \mathbb{C}^\natural)$, defines a periodicity operator

$$\mathrm{Ext}_\Lambda^n(X,Y) \ni x \to \mathbb{S}x = x \times \sigma \in \mathrm{Ext}_\Lambda^{n+2}(X,Y). \tag{15}$$

In the following we shall choose the standard generator σ that is defined 'over $\mathbb{Z}$', and then the above definition extends the periodicity operator in cyclic cohomology. This and other properties of the periodicity operator are summarized in the following Corollary ([30], Corollary 1.4)

COROLLARY. 2.3. *a) Let $x \in \mathrm{Ext}_\Lambda^n(X,Y)$ and $y \in \mathrm{Ext}_\Lambda^m(X_1,Y_1)$. Then $(\mathbb{S}x) \times y = \mathbb{S}(x \times y) = x \times (\mathbb{S}y)$.*

b) If $x \in \mathrm{Ext}_\Lambda^n(\mathbb{C}^\natural, X)$, then $\mathbb{S}x = \sigma \circ x$.

c) If $y \in \mathrm{Ext}_\Lambda^m(Y, \mathbb{C}^\natural)$, then $\mathbb{S}y = y \circ \sigma$.

d) For any extension x, we have $\mathbb{S}x = \sigma \times x$.

Using the periodicity operator, we extend the definition of periodic cyclic cohomology groups from algebras to cyclic vector spaces by

$$\mathrm{HP}^i(X) = \lim_{\to} \mathrm{Ext}_\Lambda^{i+2n}(X, \mathbb{C}^\natural), \tag{16}$$

the inductive limit being with respect to $\mathbb{S}$; clearly, $\mathrm{HP}^i(A^\natural) = \mathrm{HP}^i(A)$. Then Corollary 2.3 a) shows that the external product $\times$ is compatible with the periodicity morphism, and hence defines an external product,

$$\mathrm{HP}^i(A) \times \mathrm{HP}^j(B) \xrightarrow{\otimes} \mathrm{HP}^{i+j}(A \otimes B), \tag{17}$$

on periodic cyclic cohomology.

2.2. Extensions of algebras and products. Cyclic vector spaces will be used to study exact sequences of algebras. Let $I \subset A$ be a two–sided ideal of a complex unital algebra A (recall that in this paper all algebras are complex algebras.) Denote by $(A, I)^\natural$ the kernel of the map $A^\natural \to (A/I)^\natural$, and by $[A, I] \in \mathrm{Ext}^1_\Lambda((A/I)^\natural, (A, I)^\natural)$ the (equivalence class of the) exact sequence

$$0 \to (A, I)^\natural \to A^\natural \to (A/I)^\natural \to 0 \tag{18}$$

of cyclic vector spaces.

Let $\mathrm{HC}^i(A, I) = \mathrm{Ext}^i_\Lambda((A, I)^\natural, \mathbb{C}^\natural)$, then the long exact sequence of *Ext*–groups associated to the short exact sequence (18) reads

$$\cdots \to \mathrm{HC}^i(A/I) \to \mathrm{HC}^i(A) \to \mathrm{HC}^i(A, I) \to \mathrm{HC}^{i+1}(A/I) \to \mathrm{HC}^{i+1}(A) \to \cdots$$

By standard homological algebra, the boundary map of this long exact sequence is given by the product

$$\mathrm{HC}^i(A, I) \ni \xi \to \xi \circ [A, I] \in \mathrm{HC}^{i+1}(A/I).$$

For an arbitrary algebra I, possibly without unit, we let $I^\flat = (I^+, I)^\natural$. Then the isomorphism (14) becomes $\mathrm{HC}^n(I) \simeq \mathrm{Ext}^n_\Lambda(I^\flat, \mathbb{C}^\natural)$, and the excision theorem in periodic cyclic cohomology for cyclic vector spaces takes the following form.

Theorem. 2.4 (Cuntz–Quillen). *The inclusion $j_{I,A} : I^\flat \hookrightarrow (A, I)^\natural$ of cyclic vector spaces induces an isomorphism $\mathrm{HP}^*(A, I) \simeq \mathrm{HP}^*(I)$.*

It follows that every element $\xi \in \mathrm{HP}^*(I)$ is of the form $\xi = \xi_0 \circ j_{I,A}$, and that the boundary morphism $\partial_{A,I} : \mathrm{HP}^*(I) \to \mathrm{HP}^{*+1}(A/I)$ satisfies

$$\partial_{A,I}(\xi_0 \circ j_{I,A}) = \xi_0 \circ [A, I] \tag{19}$$

for all $\xi_0 \in \mathrm{HC}^i(A, I) = \mathrm{Ext}^i_\Lambda((A, I)^\natural, \mathbb{C}^\natural)$. Formula (19) then uniquely determines $\partial_{I,A}$.

We shall need in what follows a few properties of the isomorphisms $j_{I,A}$. Let B be an arbitrary unital algebra and I an arbitrary, possibly non–unital algebra. The inclusion $(I \otimes B)^+ \to I^+ \otimes B$, of unital algebras, defines a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & (I \otimes B)^\flat \arrow[r] \arrow[d, "\eta_{I,B}"'] & (I \otimes B)^{+\natural} \arrow[r] \arrow[d] & \mathbb{C}^\natural \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & I^\flat \times B^\natural \arrow[r] & (I^+ \otimes B)^\natural \arrow[r] & B^\natural \arrow[r] & 0
\end{tikzcd}$$

with exact lines. The morphism $\eta_{I,B}$, defined for possibly non-unital algebras I, will replace the identification $A^\natural \times B^\natural = (A \otimes B)^\natural$, valid only for unital algebras A.

Using the notation of Theorem 2.4, we see that $\eta_{I,B} = j_{I \otimes B, I^+ \otimes B}$, and hence, by the same theorem, it follows that $\eta_{I,B}$ induces an isomorphism

$$\mathrm{HP}^*(I^\flat \times B^\natural) \ni \alpha \to \alpha \circ \eta_{I,B} \in \mathrm{HP}^*(I \otimes B).$$

Using this isomorphism, we extend the external product

$$\otimes : \mathrm{HP}^*(I) \otimes \mathrm{HP}^*(B) \to \mathrm{HP}^*(I \otimes B)$$

to a possibly non-unital algebra I by

$$\begin{aligned} \mathrm{HP}^i(I)\otimes \mathrm{HP}^j(B) &= \lim_{\rightarrow} \mathrm{Ext}_\Lambda^{i+2n}(I^\flat,\mathbb{C}^\natural)\otimes \lim_{\rightarrow} \mathrm{Ext}_\Lambda^{j+2m}(B^\natural,\mathbb{C}^\natural) \\ &\xrightarrow{\times} \lim_{\rightarrow} \mathrm{Ext}_\Lambda^{i+j+2l}(I^\flat\times B^\natural,\mathbb{C}^\natural) = \mathrm{HP}^*(I^\flat\times B^\natural)\simeq \mathrm{HP}^{i+j}(I\otimes B). \end{aligned}$$

This extension of the external tensor product $\otimes$ to possibly non-unital algebras will be used to study the tensor product by B of an exact sequence $0\to I\to A\to A/I\to 0$ of algebras.

Tensoring by B is an exact functor, and hence we obtain an exact sequence

$$0\to I\otimes B\to A\otimes B\to (A/I)\otimes B\to 0. \tag{20}$$

LEMMA. 2.5. *Using the notation introduced above, we have the relation*

$$[A\otimes B, I\otimes B]=[A,I]\times \mathrm{id}_B\in \mathrm{Ext}_\Lambda^1((A/I\otimes B)^\natural,(A\otimes B,I\otimes B)^\natural).$$

Proof. We need only observe that the relation $A^\natural\times B^\natural=(A\times B)^\natural$ and the exactness of the functor $X\to X\times B^\natural$ imply that $(A,I)^\natural\times B^\natural=(A\otimes B,I\otimes B)^\natural$. □

2.3. PROPERTIES OF THE BOUNDARY MAP. The following theorem is a key tool in establishing further properties of the boundary map in periodic cyclic homology.

THEOREM. 2.6. *Let A and B be complex unital algebras and $I\subset A$ be a two-sided ideal. Then the boundary maps*

$$\partial_{I,A}:\mathrm{HP}^*(I)\to \mathrm{HP}^{*+1}(A/I)$$

and

$$\partial_{I\otimes B,A\otimes B}:\mathrm{HP}^*(I\otimes B)\to \mathrm{HP}^{*+1}((A/I)\otimes B)$$

satisfy

$$\partial_{I\otimes B,A\otimes B}(\xi\otimes\zeta)=\partial_{I,A}(\xi)\otimes\zeta$$

for all $\xi\in \mathrm{HP}^*(I)$ and $\zeta\in\mathrm{HP}^*(B)$.

Proof. The groups $\mathrm{HP}^k(I)$ is the inductive limit of the groups $\mathrm{Ext}_\Lambda^{k+2n}(I^\flat,\mathbb{C}^\natural)$ so ξ will be the image of an element in one of these Ext–groups. By abuse of notation, we shall still denote that element by ξ, and thus we may assume that $\xi\in\mathrm{Ext}_\Lambda^k(I^\flat,\mathbb{C}^\natural)$, for some large k. Similarly, we may assume that $\zeta\in\mathrm{Ext}_\Lambda^j(B^\natural,\mathbb{C}^\natural)$. Moreover, by Theorem 2.4, we may assume that $\xi=\xi_0\circ j_{I,A}$, for some $\xi_0\in\mathrm{Ext}_\Lambda^i((A,I)^\natural,\mathbb{C}^\natural)$.

We then have

$$\begin{aligned} &\partial_{I,A}(\xi)\otimes\zeta=\partial(\xi_0\circ j_{I,A})\times\zeta= && \\ &= (\xi_0\circ[A,I])\times\zeta && \text{by equation (19)}\\ &= (\mathrm{id}_{\mathbb{C}^\natural}\times\zeta)\circ((\xi_0\circ[A,I])\times\mathrm{id}_B) && \text{by Lemma 2.2}\\ &= (\mathrm{id}_{\mathbb{C}^\natural}\times\zeta)\circ(\xi_0\times\mathrm{id}_B)\circ([A,I]\times\mathrm{id}_B) && \text{by Lemma 2.2}\\ &= (\xi_0\times\zeta)\circ[A\otimes B,I\otimes B] && \text{by Lemma 2.2 and Corollary 2.3}\\ &= \partial_{A\otimes B,I\otimes B}((\xi_0\times\zeta)\circ j_{I\otimes B,A\otimes B}) && \text{by equation (19).} \end{aligned}$$

By definition, the morphism $j_{I,A}$ introduced in Theorem 2.4 satisfies

$$j_{I\otimes B,A\otimes B}=(j_{I,A}\times\mathrm{id}_B)\circ\eta_{I,B}. \tag{21}$$

Equation (21) then gives

$$\partial_{I,A}(\xi)\otimes\zeta=\partial_{I\otimes B,A\otimes B}((\xi\times\zeta)\circ\eta_{I,B})$$

in $\mathrm{Ext}_\Lambda^{i+j+1}((A/I\otimes B)^\natural,\mathbb{C}^\natural)$. This completes the proof in view of the definition of the external product $\otimes$ in the non-unital case: $\xi\otimes\zeta=(\xi\times\zeta)\circ\eta_{I,B}$. □

We now consider crossed products. Let A be a unital algebra and Γ a discrete group acting on A by $\Gamma \times A \ni (\gamma, a) \to \alpha_\gamma(a) \in A$. Then the (algebraic) crossed product $A \rtimes \Gamma$ consists of finite linear combinations of elements of the form $a\gamma$, with the product rule $(a\gamma)(b\gamma_1) = a\alpha_\gamma(b)\gamma\gamma_1$. Let $\delta(a\gamma) = a\gamma \otimes \gamma$, which defines a morphism $\delta : A \rtimes \Gamma \to A \rtimes \Gamma \otimes \mathbb{C}[\Gamma]$. Using δ, we define on $\mathrm{HP}^*(A \rtimes \Gamma)$ a $\mathrm{HP}^*(\mathbb{C}[\Gamma])$–module structure [28] by

$$\mathrm{HP}^*(A \rtimes \Gamma) \otimes \mathrm{HP}^*(\mathbb{C}[\Gamma]) \xrightarrow{\otimes} \mathrm{HP}^*((A \rtimes \Gamma) \otimes \mathbb{C}[\Gamma]) \xrightarrow{\delta^*} \mathrm{HP}^*(A \rtimes \Gamma).$$

A Γ–invariant two-sided ideal $I \subset A$ gives rise to a "crossed product exact sequence"

$$0 \to I \rtimes \Gamma \to A \rtimes \Gamma \to (A/I) \rtimes \Gamma \to 0$$

of algebras. The following theorem describes the behavior of the boundary map of this exact sequence with respect to the $\mathrm{HP}^*(\mathbb{C}[\Gamma])$–module structure on the corresponding periodic cyclic cohomology groups.

Theorem. 2.7. *Let Γ be a discrete group acting on the unital algebra A, and let I be a Γ-invariant ideal. Then the boundary map*

$$\partial_{I \rtimes \Gamma, A \rtimes \Gamma} : \mathrm{HP}^*(I \rtimes \Gamma) \to \mathrm{HP}^{*+1}((A/I) \rtimes \Gamma)$$

is $\mathrm{HP}^*(\mathbb{C}[\Gamma])$*-linear.*

Proof. The proof is based on the previous theorem, Theorem 2.6, and the naturality of the boundary morphism in periodic cyclic cohomology.

From the commutative diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & I \rtimes \Gamma & \longrightarrow & A \rtimes \Gamma & \longrightarrow & (A/I) \rtimes \Gamma & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & (I \rtimes \Gamma) \otimes \mathbb{C}[\Gamma] & \longrightarrow & (A \rtimes \Gamma) \otimes \mathbb{C}[\Gamma] & \longrightarrow & (A/I) \rtimes \Gamma \otimes \mathbb{C}[\Gamma] & \longrightarrow & 0,
\end{array}$$

we obtain that $\delta^*\partial = \partial\delta^*$ (we have omitted the subscripts). Then, for each $x \in \mathrm{HP}^*(\mathbb{C}[\Gamma])$ and $\xi \in \mathrm{HP}^*(I \rtimes \Gamma)$, we have $\xi x = \delta^*(\xi \otimes x)$, and hence, using also Theorem 2.6, we obtain

$$\partial(\xi x) = \partial(\delta^*(\xi \otimes x)) = \delta^*(\partial(\xi \otimes x)) = \delta^*((\partial\xi) \otimes x) = (\partial\xi)x\ .$$

The proof is complete. □

For the rest of this subsection it will be convenient to work with *continuous* periodic cyclic homology. Recall that this means that all algebras have compatible locally convex topologies, that we use complete projective tensor products, and that the projections $A \to A/I$ have continuous linear splittings, which implies that $A \simeq A/I \oplus I$ as locally convex vector spaces. Moreover, since the excision theorem is known only for m–algebras [13], we shall also assume that our algebras are m–algebras, that is, that their topology is generated by a family of sub-multiplicative seminorms. Slightly weaker results hold for general topological algebras and *discrete* periodic cyclic cohomology.

There is an analog of Theorem 2.7 for actions of compact Lie groups. If G is a compact Lie group acting smoothly on a complete locally convex algebra A by

$\alpha : G \times A \to A$, then the *smooth crossed product algebra* is $A \rtimes G = C^\infty(G, A)$, with the convolution product $*$,

$$f_0 * f_1(g) = \int_G f_0(h)\alpha_h(f_1(h^{-1}g))dh,$$

the integration being with respect to the normalized Haar measure on G. As before, if $I \subset A$ is a *complemented* G-invariant ideal of A, we get an exact sequence of smooth crossed products

$$0 \to I \rtimes G \to A \rtimes G \to (A/I) \rtimes G \to 0. \tag{22}$$

Still assuming that G is compact, let $R(G)$ be the representation ring of G. Then the group $\mathrm{HP}^*(A \rtimes G)$ has a natural $R(G)$–module structure defined as follows (see also [31]). The diagonal inclusion $A \rtimes G \hookrightarrow M_n(A) \rtimes G$ induces an isomorphism in cyclic cohomology, with inverse induced by the morphism

$$\frac{1}{n}Tr : M_n(A \rtimes G)^\natural \to (A \rtimes G)^\natural$$

of cyclic objects. Then, for any representation $\pi : G \to M_n(\mathbb{C})$, we obtain a unit preserving morphism

$$\mu_\pi : A \rtimes G \to M_n(A \rtimes G),$$

defined by $\mu_\pi(f)(g) = f(g)\pi(g) \in C^\infty(G, M_n(A))$, for any $f \in C^\infty(G, A)$. Finally, if $\pi \in R(G)$, we define the multiplication by π to be the morphism

$$(Tr \circ \mu_\pi)^* : \mathrm{HP}^*_{\mathrm{cont}}(A \rtimes G) \to \mathrm{HP}^*_{\mathrm{cont}}(A \rtimes G).$$

Thus, $\pi x = x \circ Tr \circ \mu_\pi$.

THEOREM. 2.8. *Let A be a locally convex m–algebra and $I \subset A$ a complemented G–invariant two-sided ideal. Then the boundary morphism associated to the exact sequence* (22),

$$\partial_{I \rtimes G, A \rtimes G} : \mathrm{HP}^*_{\mathrm{cont}}(I \rtimes G) \to \mathrm{HP}^{*+1}_{\mathrm{cont}}((A/I) \rtimes G),$$

is $R(G)$-linear.

Proof. First, we observe that the morphism $Tr : M_n(A)^\natural \to A^\natural$ is functorial, and, consequently, that it gives a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & X \arrow[r] \arrow[d] & M_n(A \rtimes G)^\natural \arrow[r] \arrow[d] & (M_n(A/I) \rtimes G)^\natural \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & (A \rtimes G, I \rtimes G)^\natural \arrow[r] & (A \rtimes G)^\natural \arrow[r] & ((A/I) \rtimes G)^\natural \arrow[r] & 0
\end{tikzcd}$$

where $X = (M_n(A \rtimes G), M_n(I \rtimes G))^\natural$ and whose vertical arrows are given by Tr.

Regarding this commutative diagram as a morphism of extensions, we obtain that

$$Tr \circ [M_n(A) \rtimes G, M_n(I) \rtimes G] = [A \rtimes G, I \rtimes G] \circ Tr. \tag{23}$$

Then, using a similar reasoning, we also obtain that

$$[M_n(A) \rtimes G, M_n(I) \rtimes G] \circ \mu_\pi = \mu_\pi \circ [A \rtimes G, I \rtimes G]. \tag{24}$$

Now let $\xi \in \mathrm{HP}^*_{\mathrm{cont}}(I \rtimes G)$, which we may assume, by Theorem 2.4, to be an element of the form $\xi = \xi_0 \circ j_{I \rtimes G, A \rtimes G}$, for some $\xi_0 \in \mathrm{Ext}^i_\Lambda((A \rtimes G, I \rtimes G)^\natural, \mathbb{C}^\natural)$. Using equations (23) and (24) and that the inclusion $j = j_{I\rtimes G, A\rtimes G}$, by the naturality of μ_π, is $R(G)$-linear, we finally get

$$\begin{aligned} &\partial(\pi\xi) = \partial(\pi(\xi_0 \circ j)) = \partial((\pi\xi_0)\circ j) = \\ = \quad &\partial(\xi_0 \circ Tr \circ \mu_\pi \circ j) = \xi_0 \circ Tr \circ \mu_\pi \circ [A \rtimes G, I \rtimes G] = \\ = \quad &\xi_0 \circ [M_n(A\rtimes G), M_n(I \rtimes G)] \circ Tr \circ \mu_\pi = \partial(\xi) \circ Tr \circ \mu_\pi = \pi\partial(\xi) \end{aligned}$$

The proof is now complete. □

In the same spirit and in the same framework as in Theorem 2.8, we now consider the action of Lie algebra cohomology on the periodic cyclic cohomology exact sequence.

Assume that G is compact and connected, and denote by $\mathfrak{g}$ its Lie algebra and by $H_*(\mathfrak{g})$ the Lie algebra homology of $\mathfrak{g}$. Since G is compact and connected, we can identify $\mathrm{H}_*(\mathfrak{g})$ with the bi-invariant currents on G. Let $\mu : G \times G \to G$ be the multiplication. Then one can alternatively define the product on $\mathrm{H}_*(\mathfrak{g})$ as the composition

$$\begin{aligned} &\mathrm{H}_*(\mathfrak{g}) \otimes \mathrm{H}_*(\mathfrak{g}) \simeq \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(G)) \otimes \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(G)) \\ &\xrightarrow{\times} \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(G\times G)) \xrightarrow{\mu^*} \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(G)) \simeq \mathrm{H}_*(\mathfrak{g}). \end{aligned}$$

We now recall the definition of the product $\mathrm{H}_*(\mathfrak{g}) \otimes \mathrm{HP}^*_{\mathrm{cont}}(A) \to \mathrm{HP}^*_{\mathrm{cont}}(A)$. Denote by $\varphi : A \to C^\infty(G, A)$ the morphism $\varphi(a)(g) = \alpha_g(a)$, where, this time, $C^\infty(G,A)$ is endowed with the *pointwise* product. Then $x \times \xi \in \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(G)\widehat{\otimes}A)$ is a (continuous) cocycle on $C^\infty(G,A) \simeq C^\infty(G)\widehat{\otimes}A$, and we define $x\xi = \varphi^*(x\otimes\xi)$. The associativity of the $\times$-product shows that $\mathrm{HP}^*_{\mathrm{cont}}(A)$ becomes a $\mathrm{H}_*(\mathfrak{g})$–module with respect to this action.

Theorem. 2.9. *Suppose that a compact connected Lie group G acts smoothly on a complete locally convex algebra A and that I is a closed invariant two-sided ideal of A, complemented as a topological vector space. Then*

$$\partial(x\xi) = x(\partial\xi),$$

for any $x \in \mathrm{H}_*(\mathfrak{g})$ and $\xi \in \mathrm{HP}^*_{\mathrm{cont}}(I)$.

Proof. The proof is similar to the proof of Theorem 2.8, using the morphism of exact sequences

$$\begin{tikzcd} 0 \arrow[r] & (A,I)^\natural \arrow[r] \arrow[d] & A^\natural \arrow[r] \arrow[d] & (A/I)^\natural \arrow[r] \arrow[d] & 0 \\ 0 \arrow[r] & X \arrow[r] & C^\infty(G,A)^\natural \arrow[r] & C^\infty(G,A/I)^\natural \arrow[r] & 0 \end{tikzcd}$$

where $X = (C^\infty(G,A), C^\infty(G,I))^\natural$. □

2.4. Relation to the bivariant Chern–Connes character. A different type of property of the boundary morphism in periodic cyclic cohomology is its compatibility (effectively an identification) with the bivariant Chern-Connes character [30]. Before we can state this result, need to recall a few constructions from [30].

Let A and B be unital locally convex algebras and assume that a continuous linear map

$$\beta : A \to \mathcal{B}(\mathcal{H})\widehat{\otimes}B$$

is given, such that the cocycle $\ell(a_0, a_1) = \beta(a_0)\beta(a_1) - \beta(a_0a_1)$ factors as a composition $A\widehat{\otimes}A \to \mathcal{C}_p(\mathcal{H})\widehat{\otimes}B \to \mathcal{B}(\mathcal{H})\widehat{\otimes}B$ of continuous maps. (Recall that $\mathcal{C}_p(\mathcal{H})$ is the ideal of p–summable operators and that $\widehat{\otimes}$ is the complete projective tensor product.) Using the cocycle ℓ, we define on $E_\beta = A \oplus \mathcal{C}_p(\mathcal{H})\hat{\otimes}B$ an associative product by the formula

$$(a_1, x_1)(a_2, x_2) = (a_1a_2, \beta(a_1)x_2 + x_1\beta(a_2) + \ell(a_1, a_2)).$$

Then the algebra E_β fits into the exact sequence

$$0 \to \mathcal{C}_p(\mathcal{H})\widehat{\otimes}B \to E_\beta \to A \to 0. \tag{25}$$

An exact sequence

$$[E]: \quad 0 \to \mathcal{C}_p(\mathcal{H})\widehat{\otimes}B \to E \to A \to 0. \tag{26}$$

that is isomorphic to an exact sequence of the form (25) will be called an *admissible* exact sequence. If $[E]$ is an admissible exact sequence and $n \geq p-1$, then [30, Theorem 3.5] associates to $[E]$ an element

$$ch_1^{2n+1}([E]) \in \mathrm{Ext}^{2n+1}_{\Lambda,\mathrm{cont}}(A^\natural, B^\natural), \tag{27}$$

which for $B = \mathbb{C}$ recovers Connes' Chern character in K-homology [10]. (The subscript "cont" stresses that we are considering the version of the Yoneda Ext defined for locally convex cyclic objects.)

Let $Tr : \mathcal{C}_1(\mathcal{H}) \to \mathbb{C}$ be the ordinary trace, i.e., $Tr(T) = \sum_n (Te_n, e_n)$ for any orthonormal basis $(e_n)_{n\geq 0}$ of the Hilbert space $\mathcal{H}$. Using the trace Tr we define $Tr_n \in \mathrm{HC}^{2n}(\mathcal{C}_p(\mathcal{H}))$, for $2n \geq p-1$, to be the class of the cyclic cocycle

$$Tr_n(a_0, a_1, \ldots, a_{2n}) = (-1)^n \frac{n!}{(2n)!} Tr(a_0a_1 \ldots a_{2n}). \tag{28}$$

The normalization factor was chosen such that $Tr_n = \mathbb{S}^n Tr_1 = \mathbb{S}^n Tr$ on $C_1(\mathcal{H})$. We have the following compatibility between the bivariant Chern-Connes character and the Cuntz–Quillen boundary morphism.

Let $\mathrm{HP}^*_{\mathrm{cont}} \ni \xi \to \xi_{\mathrm{disc}} \in \mathrm{HP}^*_{\mathrm{disc}} := \mathrm{HP}^*$ be the natural transformation that "forgets continuity" from continuous to ordinary (or discrete) periodic cyclic cohomology. We include the subscript "disc" only when we need to stress that discrete homology is used. By contrast, the subscript "cont" will always be included.

THEOREM. 2.10. *Let $0 \to \mathcal{C}_p(\mathcal{H})\widehat{\otimes}B \to E \to A \to 0$ be an admissible exact sequence and $ch_1^{2n+1}([E]) \in \mathrm{Ext}^{2n+1}_{\Lambda,\mathrm{cont}}(A^\natural, B^\natural)$ be its bivariant Chern–Connes character, equation* (27)*. If Tr_n is as in equation* (28) *and $n \geq p-1$, then*

$$\partial(Tr_n \otimes \xi)_{\mathrm{disc}} = (\xi \circ ch_1^{2n+1}([E]))_{\mathrm{disc}} \in \mathrm{HP}^{q+1}(A),$$

for each $\xi \in \mathrm{HP}^q_{\mathrm{cont}}(B)$.

This theorem provides us–at least in principle–with formulæ to compute the boundary morphism in periodic cyclic cohomology, see [29] and [30], Proposition 2.3.

Before proceeding with the proof, we recall a construction implicit in [30]. The algebra $RA = \oplus_{j\geq 0} A^{\hat{\otimes} j}$ is the tensor algebra of A, and rA is the kernel of the map $RA \to A^+$. Because A has a unit, we have a canonical isomorphism $A^+ \simeq \mathbb{C} \oplus A$. We do not consider any topology on RA, but in addition to $(RA)^\natural$, the cyclic object associated to RA, we consider a completion of it in a natural topology with respect to which all structural maps are continuous. The new, completed, cyclic object is denoted $(RA)^{\natural}_{\rm cont}$ and is obtained as follows. Let $R_kA = \oplus_{j=0}^{k} A^{\hat{\otimes} j}$. Then

$$(RA)^{\natural}_{{\rm cont},n} = \lim_{k\to\infty} (R_kA)^{\hat{\otimes} n+1},$$

with the inductive limit topology.

Proof. We begin with a series of reductions that reduce the proof of the Theorem to the proof of (29).

Since $[E]$ is an admissible extension, there exists by definition a continuous linear section $s : A \to E$ of the projection $\pi : E \to A$ (i.e., $\pi \circ s = id$). Then s defines a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & rA & \longrightarrow & RA & \longrightarrow & A^+ & \longrightarrow & 0 \\ & & \Big\downarrow{\scriptstyle\varphi} & & \Big\downarrow{\scriptstyle\psi} & & \Big\downarrow{\scriptstyle\pi_A} & & \\ 0 & \longrightarrow & \mathcal{C}_p\widehat{\otimes}B & \longrightarrow & E & \longrightarrow & A & \longrightarrow & 0\ , \end{array}$$

where the right hand vertical map is the projection $A^+ \simeq \mathbb{C}\oplus A \to A$.

By increasing q if necessary, we may assume that the cocycle $\xi \in {\rm HP}^q_{\rm cont}(B)$ comes from a cocycle, also denoted ξ, in ${\rm HC}^q_{\rm cont}(B)$. Let

$$\xi_1 = (Tr_n \otimes \xi)_{\rm disc} \in {\rm HC}^{q+2n}_{\rm disc}(\mathcal{C}_p\widehat{\otimes}B) := {\rm HC}^{q+2n}(\mathcal{C}_p\widehat{\otimes}B)$$

be as in the statement of the theorem.

We claim that it is enough to show that

$$\partial(\varphi^*\xi_1)\circ j_A = (\xi \circ ch_1^{2n+1}([E]))_{\rm disc}, \tag{29}$$

where $j_A = A^{\natural} \to (A^+)^{\natural}$ is the inclusion.

Indeed, assuming (29) and using the above commutative diagram and the naturality of the boundary morphism, we obtain

$$(\xi \circ ch_1^{2n+1}([E]))_{\rm disc} = \partial(\varphi^*\xi_1)\circ j_A = \pi_A^*(\partial\xi_1)\circ j_A = \partial\xi_1 \circ \pi_A \circ j_A = \partial\xi_1,$$

as stated in theorem, because $\pi_A \circ j_A = id$.

Let $j_{rA,RA} : (rA)^{\flat} \hookrightarrow (RA, rA)^{\natural}$ be the morphism (inclusion) considered in Theorem 2.4. Also, let $\xi_2 \in {\rm HC}^n_{\rm disc}((RA,rA)^{\natural}) = {\rm Ext}^n_{\Lambda}((RA,rA)^{\natural}, \mathbb{C}^{\natural})$ satisfy

$$\xi_2 \circ j_{rA,RA} = \varphi^*\xi_1 \in {\rm HC}^n_{\rm disc}((rA)^{\flat}) = {\rm Ext}^n_{\Lambda}((rA)^{\flat}, \mathbb{C}^{\natural}). \tag{30}$$

(In words: "ξ_2 restricts to $\varphi^*\xi_1$ on $(rA)^{\flat}$.") Then, using equation (19), we have

$$\partial(\varphi^*\xi_1) = \xi_2 \circ [RA, rA]. \tag{31}$$

The rest of the proof consists of showing that the construction of the odd bivariant Chern-Connes character [30] provides us with ξ_2 satisfying equations (30) and (32):

$$\xi_2 \circ [RA, rA]\circ j_A = (\xi \circ ch_1^{2n+1}([E]))_{\rm disc}. \tag{32}$$

This is enough to complete the proof because equations (31) and (32) imply (29) and, as we have already shown, equation (30) implies equation (31). So, to complete the proof, we now proceed to construct ξ_2 satisfying (30) and (32).

Recall from [30] that the ideal rA defines a natural increasing filtration of $(RA)^{\natural}_{\rm cont}$ by cyclic vector spaces:

$$(RA)^{\natural}_{\rm cont} = F_0(RA)^{\natural}_{\rm cont} \supset F_{-1}(RA)^{\natural}_{\rm cont} \supset \dots,$$

such that $(rA)^\flat \subset F_{-1}(RA)^\natural_{\text{cont}} = (RA, rA)^\natural$. If $(rA)^\flat_k$ is the k–th component of the cyclic vector space $(rA)^\flat$ (and if, in general, the lower index stands for the $\mathbb{Z}_+$–grading of a cyclic vector space) then we have the more precise relation

$$(rA)^\flat_k \subset (F_{-n-1}(RA)^\natural_{\text{cont}})_k \ , \qquad \text{for } k \geq n. \tag{33}$$

It follows that the morphism of cyclic vector spaces

$$\tilde{\tau}_n = Tr \circ F_{-n-1}(\psi) : F_{-n-1}(RA)^\natural_{\text{cont}} \to B^\natural$$

(defined in [30], page 579) satisfies $\tilde{\tau}_n = Tr \circ \varphi$ on $(rA)^\flat_k$, for $k \geq n \geq p-1$. Fix then $k = q+2n$, and conclude that $\xi_1 = Tr_n \otimes \xi_{\text{disc}} \in \mathrm{HC}^{q+2n}(\mathcal{C}_p \widehat{\otimes} B)$ satisfies

$$\varphi^* \xi_1 = \varphi^*(Tr_n \otimes \xi) = \xi_{\text{disc}} \circ \mathbb{S}^n \tilde{\tau}_n \tag{34}$$

on $(rA)^\flat_k \subset F_{-n-1}(RA)^\natural_{\text{cont}}$, because Tr_n restricts to $\mathbb{S}^n Tr$ on $\mathcal{C}_1(\mathcal{H})$. Now recall the crucial fact that there exists an extension

$$C_0^{2n}(RA) \in \mathrm{Ext}^{2n}_{\Lambda,\text{cont}}(F_{-1}(RA)^\natural_{\text{cont}}, F_{-n-1}(RA)^\natural_{\text{cont}})$$

that has the property that $C_0^{2n}(RA) \circ i = \mathbb{S}^n$, if $i : F_{-n-1}(RA)^\natural_{\text{cont}} \to F_{-1}(RA)^\natural_{\text{cont}}$ is the inclusion (see [30], Corollary 2.2). Using this extension, we finally define

$$\xi_2 = (\xi \circ \tilde{\tau}_n \circ C_0^{2n}(RA))_{\text{disc}} \in \mathrm{Ext}^n_\Lambda(F_{-1}(RA)^\natural_{\text{cont}}, \mathbb{C}^\natural).$$

Since ξ_2 has order $k = q+2n \geq 2n \geq n$, we obtain from the equations (33) and (34) that ξ_2 satisfies (30) (i.e., that it restricts to $\varphi^*\xi_1$ on $(rA)^\flat_k \subset F_{-n-1}(RA)^\natural_{\text{cont}}$), as desired.

The last thing that needs to be checked for the proof to be complete is that ξ_2 satisfies equation (32). By definition, the odd bivariant Chern-Connes character ([30], page 579) is

$$ch_1^{2n+1}([E]) = \tilde{\tau}_n \circ ch_1^{2n+1}(RA) \circ j_A, \tag{35}$$

where $ch_1^{2n+1}(RA) = C_1^{2n+1}(RA) = C_0^{2n}(RA) \circ q_0(RA)$, and $j_A : A^\natural \to (A^+)^\natural$ is the inclusion (see [30], page 568, definition 2.4. page 574, and the discussion on page 579). Moreover $q_0(RA)$ is nothing but a continuous version of $[RA, rA]$, that is

$$q_0(RA)_{\text{disc}} = [RA, rA],$$

and hence

$$\xi_2 \circ [RA, rA] \circ j_A = (\xi \circ \tilde{\tau}_n \circ C_0^{2n}(RA) \circ q_0(RA) \circ j_A)_{\text{disc}} = (\xi \circ ch_1^{2n+1}([E]))_{\text{disc}}.$$

Since ξ_2 satisfies equation (30) and (32), which imply equation (29), the proof is complete. $\square$

For any locally convex algebra B and $\xi \in \mathrm{HP}^*(B)$, the *discrete* periodic cyclic cohomology of B, we say that ξ is a *continuous class* if it can be represented by a continuous cocycle on B. Put differently, this means that $\xi = \zeta_{\text{disc}}$, for some $\zeta \in \mathrm{HP}^*_{\text{cont}}(B)$. Since the bivariant Chern–Connes character can, at least in principle, be expressed by an explicit formula, it preserves continuity. This gives the following corollary.

COROLLARY. 2.11. *The periodic cyclic cohomology boundary map ∂ associated to an admissible extension maps a class of the form $Tr_n \otimes \xi$, for ξ a continuous class, to a continuous class.*

It is likely that recent results of Cuntz, see [12, 13], will give the above result for all continuous classes in $\mathrm{HP}^*(\mathcal{C}_p\hat{\otimes}B)$ (not just the ones of the form $Tr_n \otimes \xi$).

Using the above corollary, we obtain the compatibility between the bivariant Chern–Connes character and the index morphism in full generality. This result had been known before only in particular cases [30].

Theorem. 2.12. *Let $0 \to \mathcal{C}_p(\mathcal{H})\widehat{\otimes}B \to E \to A \to 0$ be an admissible exact sequence and $ch_1^{2n+1}([E]) \in \mathrm{Ext}_\Lambda^{2n+1}(A^\natural, B^\natural)$ be its bivariant Chern–Connes character, equation (27). If Tr_n is as in equation (28) and $\mathrm{Ind} : \mathrm{K}_1^{\mathrm{alg}}(A) \to \mathrm{K}_0^{\mathrm{alg}}(\mathcal{C}_p(\mathcal{H})\widehat{\otimes}B)$ is the connecting morphism in algebraic K–Theory then, for any $\varphi \in \mathrm{HP}^0_{\mathrm{cont}}(B)$ and $[u] \in \mathrm{K}_1^{\mathrm{alg}}(A)$, we have*

$$\langle\, Tr_n \otimes \varphi, \mathrm{Ind}[u]\,\rangle = \langle\, ch_1^{2n+1}([E]) \circ \varphi, [u]\,\rangle\,. \tag{36}$$

3. The index theorem for coverings

Using the methods we have developed, we now give a new proof of Connes–Moscovici's index theorem for coverings. To a covering $\widetilde{M} \to M$ with covering group Γ, Connes and Moscovici associated an extension

$$0 \longrightarrow \mathcal{C}_{n+1} \otimes \mathbb{C}[\Gamma] \longrightarrow E_{CM} \longrightarrow C^\infty(S^*M) \longrightarrow 0, \quad n = \dim M,$$

(the Connes–Moscovici exact sequence), defined using invariant pseudodifferential operators on $\widetilde{M}$; see equation (45). If $\varphi \in \mathrm{H}^*(\Gamma) \subset \mathrm{HP}^*_{\mathrm{cont}}(\mathcal{C}_{n+1} \otimes \mathbb{C}[\Gamma])$ is an even cyclic cocycle, then the Connes–Moscovici index theorem computes the morphisms

$$\varphi_* \circ \mathrm{Ind} : \mathrm{K}_1^{\mathrm{alg}}(C^\infty(S^*M)) \longrightarrow \mathbb{C},$$

where Ind is the index morphism associated to the Connes–Moscovici exact sequence. Our method of proof then is to use the compatibility between the connecting morphisms in algebraic K–Theory and ∂, the connecting morphism in periodic cyclic cohomology (Theorem 1.5), to reduce the proof to the computation of ∂. This computation is now a problem to which the properties of ∂ established in Section 2 can be applied.

We first show how to obtain the Connes–Moscovici exact sequence from another exact sequence, the Atiyah–Singer exact sequence, by a purely algebraic construction. Then, using the naturality of ∂ and Theorem 2.6, we determine the connecting morphism ∂_{CM} of the Connes–Moscovici exact sequence in terms of the connecting morphism ∂_{AS} of the Atiyah–Singer exact sequence. For the Atiyah–Singer exact sequence the procedure can be reversed and we now use the Atiyah-Singer Index Theorem and Theorem 1.5 to compute ∂_{AS}.

A comment about the interplay of continuous and discrete periodic cyclic cohomology in the proof below is in order. We have to use continuous periodic cyclic cohomology whenever we want explicit computations with the periodic cyclic cohomology of groupoid algebras, because only the continuous version of periodic cyclic cohomology is known for groupoid algebras associated to étale groupoids [7]. On the other hand, in order to be able to use Theorem 1.5, we have to consider ordinary (or discrete) periodic cyclic cohomology as well. This is not an essential difficulty because, using Corollary 2.11, we know that the index classes are represented by continuous cocycles.

3.1. GROUPOIDS AND THE CYCLIC COHOMOLOGY OF THEIR ALGEBRAS. Our computations are based on groupoids, so we first recall a few facts about groupoids.

A *groupoid* is a small category in which every morphism is invertible. (Think of a groupoid as a set of points joined arrows; the following examples should clarify this abstract definition of groupoids.) A *smooth étale* groupoid is a groupoid whose set of morphisms (also called *arrows*) and whose set of objects (also called *units*) are smooth manifolds such that the domain and range maps are étale (i.e., local diffeomorphisms). To any smooth étale groupoid $\mathcal{G}$, assumed Hausdorff for simplicity, there is associated the algebra $C_c^\infty(\mathcal{G})$ of compactly supported functions on the set of arrows of $\mathcal{G}$ and endowed with the convolution product $*$,

$$(f_0 * f_1)(g) = \sum_{r(\gamma)=r(g)} f_0(\gamma) f_1(\gamma^{-1} g).$$

Here r is the range map and $r(\gamma) = r(g)$ is the condition that γ^{-1} and g be composable. Whenever dealing with $C_c^\infty(\mathcal{G})$, we will use *continuous* cyclic cohomology, as in [7]. See [7] for more details on étale groupoids, and [35] for the general theory of locally compact groupoids.

Étale groupoids conveniently accommodate in the same framework smooth manifolds and (discrete) groups, two extreme examples in the following sense: the smooth étale groupoid associated to a smooth manifold M has only identity morphisms, whereas the smooth étale groupoid associated to the (discrete) group Γ has only one object, the identity of Γ. The algebras $C_c^\infty(\mathcal{G})$ associated to these groupoids are $C_c^\infty(M)$ and, respectively, the group algebra $\mathbb{C}[\Gamma]$. Here are other examples used in the paper.

The groupoid R_I associated to an equivalence relation on a discrete set I has I as the set of units and exactly one arrow for any ordered pair of equivalent objects. If I is a finite set with k elements and all objects of I are equivalent (i.e., if R_I is the *total equivalence relation* on I) then $C_c^\infty(R_I) \simeq M_k(\mathbb{C})$ and its classifying space in the sense of Grothendieck [34], the space $\mathrm{B}\, R_I$, is contractable [17, 34].

Another example, the *gluing groupoid* $\mathcal{G}_\mathcal{U}$, mimics the definition a manifold M in terms of "gluing coordinate charts." The groupoid $\mathcal{G}_\mathcal{U}$ is defined [7] using an open cover $\mathcal{U} = (U_\alpha)_{\alpha \in I}$ of M, i.e., $M = \cup_{\alpha \in I} U_\alpha$. Then $\mathcal{G}_\mathcal{U}$ has units $\mathcal{G}_\mathcal{U}^0 = \cup_{\alpha \in I} U_\alpha \times \{\alpha\}$ and arrows

$$\mathcal{G}_\mathcal{U}^{(1)} = \{(x, \alpha, \beta), \alpha, \beta \in I, x \in U_\alpha \cap U_\beta\}.$$

If R_I is the total equivalence relation on I, then there is an injective morphism $l : \mathcal{G}_\mathcal{U} \hookrightarrow M \times R_I$ of étale groupoids.

Let $f : \mathcal{G}_1 \to \mathcal{G}_2$ be an *étale morphism of groupoids*, that is, a morphism of étale groupoids that is a local diffeomorphism. Then the map f defines a continuous map, $Bf : B\mathcal{G}_2 \to B\mathcal{G}_1$, of classifying spaces and a group morphism, $f_{Tr} : \mathrm{HP}^*_{\mathrm{cont}}(C_c^\infty(\mathcal{G}_1)) \to \mathrm{HP}^*_{\mathrm{cont}}(C_c^\infty(\mathcal{G}_2))$. If f is *injective* when restricted to units, then there exists an algebra morphism $\iota(f) : C_c^\infty(\mathcal{G}_1) \to C_c^\infty(\mathcal{G}_2)$ such that $f_{TR} = \iota(f)^*$.

The following theorem, a generalization of [7], Theorem 5.7. (2), is based on the fact that all isomorphisms in the proof of that theorem are functorial with respect to étale morphisms. It is the reason why we use continuous periodic cyclic cohomology when working with groupoid algebras. Note that the cyclic object associated to $C_c^\infty(\mathcal{G})$, for G an étale groupoid, is an inductive limit of locally convex nuclear spaces.

THEOREM. 3.1. *If $\mathcal{G}$ is a Hausdorff étale groupoid of dimension n, and if $\mathfrak{o}$ is the complexified orientation sheaf of $\mathrm{B}\,\mathcal{G}$, then there exists a natural embedding $\Phi : \mathrm{H}^{*+n}(\mathrm{B}\,\mathcal{G},\mathfrak{o}) \hookrightarrow \mathrm{HP}^*_{\mathrm{cont}}(C_c^\infty(\mathcal{G}))$. Here "natural" means that if $f : \mathcal{G}_1 \to \mathcal{G}_2$ is an étale morphism of groupoids, then the diagram*

$$\begin{tikzcd}
\mathrm{H}^{*+n}(\mathrm{B}\,\mathcal{G}_2,\mathfrak{o}_2) \arrow[r] \arrow[d, "(\mathrm{B}\,f)^*"'] & \mathrm{HP}^*_{\mathrm{cont}}(C_c^\infty(\mathcal{G}_2)) \arrow[d, "f_{Tr}"] \\
\mathrm{H}^{*+n}(\mathrm{B}\,\mathcal{G}_1,\mathfrak{o}_1) \arrow[r] & \mathrm{HP}^*_{\mathrm{cont}}(C_c^\infty(\mathcal{G}_1)),
\end{tikzcd}$$

whose horizontal lines are the morphisms Φ, commutes.

For discrete groups, Theorem 3.1 recovers the embedding

$$\mathrm{H}^*(\Gamma) = \mathrm{H}^*(\mathrm{B}\,\Gamma,\mathbb{C}) \hookrightarrow \mathrm{HP}^*_{\mathrm{cont}}(\mathbb{C}[\Gamma])$$

of [8, 20].

For smooth manifolds, the embedding Φ of Theorem 3.1 is just the Poincaré duality–an isomorphism. This isomorphism has a very concrete form. Indeed, let $\xi \in \mathrm{H}^{n-i}(M,\mathfrak{o})$ be an element of the singular cohomology of M with coefficients in the orientation sheaf, let $\eta \in \mathrm{H}^i_c(M)$ be an element of the singular cohomology of M with compact supports (all cohomology groups have complex coefficients), and let

$$\chi : \mathrm{HP}_i^{\mathrm{cont}}(C_c^\infty(M)) \simeq \oplus_k \mathrm{H}^{i+2k}_{c,DR}(M) = \oplus_k \mathrm{H}^{i+2k}_c(M)$$

be the canonical isomorphism induced by the Hochschild-Kostant-Rosenberg map χ, equation (12). Then the isomorphism Φ is determined by

$$\langle \Phi(\xi), \eta\rangle = \langle \xi \wedge \chi(\eta), [M]\rangle \in \mathbb{C}, \tag{37}$$

where the first pairing is the map $\mathrm{HP}^*_{\mathrm{cont}}(C_c^\infty(M)) \otimes \mathrm{HP}_*^{\mathrm{cont}}(C_c^\infty(M)) \to \mathbb{C}$ and the second pairing is the evaluation on the fundamental class.

Typically, we shall use these results for the manifold S^*M, for which there is an isomorphism $\mathrm{H}^{*-1}(S^*M) \simeq \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(S^*M))$, because S^*M is orient<u>ed</u>. (The orientation of S^*M is the one induced from that of T^*M as in [5]. More precisely B^*M, the disk bundle of M, is given the orientation in which the "the horizontal part is real and the vertical part is imaginary," and S^*M is oriented as the boundary of an oriented manifold.) The shift in the $\mathbb{Z}_2$-degree is due to the fact that S^*M is odd dimensional.

3.2. MORITA INVARIANCE AND COVERINGS. Let M be a smooth compact manifold and $q : \widetilde{M} \to M$ be a covering with Galois group Γ; said differently, $\widetilde{M}$ is a principal Γ–bundle over M. We fix a finite cover $\mathcal{U} = (U_\alpha)_{\alpha\in I}$ of M by trivializing open sets, i.e., $q^{-1}(U_\alpha) \simeq U_\alpha \times \Gamma$ and $M = \cup U_\alpha$. The transition functions between two trivializing isomorphisms on their common domain, the open set $U_\alpha \cap U_\beta$, defines a 1–cocycle $\gamma_{\alpha\beta}$ that completely determines the covering $q : \widetilde{M} \to M$.

In what follows, we shall need to lift the covering $q : \widetilde{M} \to M$ to a covering $q : S^*\widetilde{M} \to S^*M$, using the canonical projection $p : S^*M \to M$. All constructions then lift, from M to S^*M, canonically. In particular, $V_\alpha = p^{-1}(U_\alpha)$ is a finite covering of S^*M with trivializing open sets, and the associated 1–cocycle is (still) $\gamma_{\alpha\beta}$. Moreover, if $f_0 : M \to \mathrm{B}\,\Gamma$ classifies the covering $q : \widetilde{M} \to M$, then $f = f_0 \circ p$ classifies the covering $S^*\widetilde{M} \to S^*M$.

Suppose that the trivializing cover $\mathcal{V} = (V_\alpha)_{\alpha \in I}$ of S^*M consists of k open sets, and let $\sum \varphi_\alpha^2 = 1$ be a partition of unity subordinated to $\mathcal{V}$. The cocycle identity $\gamma_{\alpha\beta}\gamma_{\beta\delta} = \gamma_{\alpha\delta}$ ensures then that the matrix

$$p = [\varphi_\alpha \gamma_{\alpha\beta} \varphi_\beta]_{\alpha,\beta \in I} \in M_k(C^\infty(M)) \otimes \mathbb{C}[\Gamma] \tag{38}$$

is an idempotent, called *the Mishchenko idempotent*; a different choice of a trivializing cover and of a partition of unity gives an equivalent idempotent.

Using the Mishchenko idempotent p, we now define the morphism

$$\lambda : C^\infty(S^*M) \to M_k(C^\infty(S^*M)) \otimes \mathbb{C}[\Gamma]$$

by $\lambda(a) = ap$, for $a \in C^\infty(S^*M)$; explicitly,

$$\lambda(a)(x) = a(x)p(x) = [a(x)\varphi_\alpha(x)\varphi_\beta(x) \otimes \gamma_{\alpha\beta}]. \tag{39}$$

Because the morphism λ is used to define the Connes–Moscovici extension, equation (45) below, we need to identify the induced morphism

$$\lambda^* : \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(S^*M) \otimes \mathbb{C}[\Gamma]) \to \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(S^*M)).$$

The identification of λ, Proposition 3.3, is based on writing λ as a composition of three simpler morphisms, morphisms that will play an auxiliary role. The next few paragraphs before Proposition 3.3 will deal with the definition and properties of these morphisms.

We define the first auxiliary morphism $\iota(g)$ to be induced by an étale morphism of groupoids. Let $\mathcal{G}_\mathcal{V}$ be the gluing groupoid associated to the cover $\mathcal{V} = (V_\alpha)_{\alpha \in I}$ of S^*M. Using the cocycle $(\gamma_{\alpha\beta})_{\alpha,\beta \in I}$ associated to $\mathcal{V}$ that identifies the covering $S^*\widetilde{M} \to S^*M$, we define the étale morphism of groupoids g by

$$\mathcal{G}_\mathcal{V} \ni (x, \alpha, \beta) \xrightarrow{g} (x, \alpha, \beta, \gamma_{\alpha\beta}) \in \mathcal{G}_\mathcal{V} \times \Gamma,$$

which induces a morphism $\iota(g) : C_c^\infty(\mathcal{G}_\mathcal{V}) \to C_c^\infty(\mathcal{G}_\mathcal{V}) \otimes \mathbb{C}[\Gamma]$ and a continuous map $\mathrm{B}\,g : \mathrm{B}\,\mathcal{G}_\mathcal{V} \to \mathrm{B}(\mathcal{G}_\mathcal{V} \times \Gamma) = \mathrm{B}\,\mathcal{G}_\mathcal{V} \times \mathrm{B}\,\Gamma$.

The projection $t : \mathcal{G}_\mathcal{V} \to S^*M$ is an etale morphism of groupoids that induces a homotopy equivalence $\mathrm{B}\,\mathcal{G}_\mathcal{V} \to S^*M$ and hence also an isomorphism

$$t_{Tr} : \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(S^*M)) \to \mathrm{HP}^*_{\mathrm{cont}}(C_c^\infty(\mathcal{G}_\mathcal{V})).$$

By definition, $t_{Tr} = Tr \circ \iota(l)^*$, where $l : \mathcal{G}_\mathcal{V} \to S^*M \times R_I$ is the natural inclusion considered also before, and Tr is the generic notation for the isomorphisms $Tr : \mathrm{HP}_*(M_n(A)) \simeq \mathrm{HP}_*(A)$, induced by the trace. In particular, $\iota(l)^*$ is also an isomorphism.

Using the homotopy equivalence $\mathrm{B}\,t$ of $\mathrm{B}\,\mathcal{G}_\mathcal{V}$ and S^*M, we obtain a continuous map

$$h_0 : S^*M \to S^*M \times \mathrm{B}\,\Gamma,$$

uniquely determined by the condition $h_0 \circ \mathrm{B}\,t = (\mathrm{B}\,t \times id) \circ \mathrm{B}\,g$.

LEMMA. 3.2. *The map h_0 defined above coincides, up to homotopy, with the product function (id_{S^*M}, f), where $f : S^*M \to \mathrm{B}\,\Gamma$ classifies $S^*\widetilde{M} \to S^*M$.*

Proof. Denote by p_1 and p_2 the projections of $S^*M \times \mathrm{B}\,\Gamma$ onto components. The map $p_1 \circ h_0$ is easily seen to be the identity, so $h_0 = \mathrm{id}_{S*M} \times h_1$ where $h_1 : S^*M \to \mathrm{B}\,\Gamma$ is induced by the *non*-étale morphism of topological groupoids $\mathcal{G}_\mathcal{V} \ni (x, \alpha, \beta) \to \gamma_{\alpha\beta} \in \Gamma$. In order to show that h_1 coincides with f, up to homotopy, it is enough to show that the principal Γ-bundle (i.e., covering) that h_1 pulls back from $\mathrm{B}\,\Gamma$ to S^*M is isomorphic to the covering $S^*\widetilde{M} \to \widetilde{M}$.

Let $\mathcal{G}_{\mathcal{U}}$ be the gluing groupoid associated to the cover $\mathcal{U} = (U_\alpha)_{\alpha \in I}$ of M. It is seen from the definition that $\mathcal{G}_{\mathcal{V}} \to \Gamma$ factors as $\mathcal{G}_{\mathcal{V}} \to \mathcal{G}_{\mathcal{U}} \to \Gamma$, where the function $\mathcal{G}_{\mathcal{V}} \to \Gamma$ acts as $(m, \alpha, \beta) \to \gamma_{\alpha\beta}$. Thus we may replace S^*M by M everywhere in the proof.

Since the the covering $\widetilde{M} \to M$ is determined by its restriction to loops, we may assume that M is the circle S^1. Cover $M = S^1$ by two contractable intervals $I_0 \cap I_1$ which intersect in two small disjoint neighborhoods of 1 and -1: $I_0 \cap I_1 = (z, z^{-1}) \cup (-z, -z^{-1})$ where $z \in S'$ and $|z-1|$ is very small. We may also assume that the transition cocycle is the identity on (z, z^{-1}) and $\gamma \in \Gamma$ on $(-z, -z^{-1})$ (we have replaced constant Γ–cocycles with locally constant Γ–cocycles). The map h_1 maps each of the units of $\mathcal{G}_{\mathcal{U}}$ and each of the 1-cells corresponding to the right hand interval (z, z^{-1}) to the only 0-cell of $\mathrm{B}\,\Gamma$, the cell corresponding to the identity $e \in \Gamma$. (Recall that the classifying space of a topological groupoid is the geometrical realization of the simplicial space of composable arrows [34], and that that there is a 0 cell for each unit, a 1-cell for each non-identity arrow, a 2-cell for each pair of non-identity composable arrows, and so on). The other 1-cells (i.e., corresponding to the arrows leaving from a point on the left hand side interval) will map to the 1-cell corresponding γ. This shows that, on homotopy groups, the induced map $\mathbb{Z} = \pi_1(S^1) \to \Gamma = \pi_1(B\Gamma)$ sends the generator 1 to γ. This completes the proof of the lemma. □

We need to introduce one more auxiliary morphism before we can determine λ^*. Using the partition of unity $\sum_\alpha \varphi_\alpha^2 = 1$ subordinated to $\mathcal{V} = (V_\alpha)_{\alpha \in I}$, we define $\nu : C^\infty(S^*M) \to C_c^\infty(\mathcal{G}_{\mathcal{V}})$ by

$$\nu(f)(x, \alpha, \beta) = f(x)\varphi_\alpha(x)\varphi_\beta(x),$$

which turns out to be a morphism of algebras. Because the composition

$$C^\infty(S^*M) \xrightarrow{\nu} C_c^\infty(\mathcal{G}_{\mathcal{V}}) \xrightarrow{\iota(l)} C_c^\infty(S^*M \times R_I) = M_k(C^\infty(S^*M))$$

is (unitarily equivalent to) the upper–left corner embedding, we obtain that the morphism $\nu^* : \mathrm{HP}^*_{\mathrm{cont}}(C_c^\infty(\mathcal{G}_{\mathcal{V}})) \to \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(S^*M))$ is the inverse of t_{Tr}.

We are now ready to determine the morphism

$$\lambda^* : \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(S^*M) \otimes \mathbb{C}[\Gamma]) \to \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(S^*M)).$$

In order to simplify notation, in the statement of the following result we shall identify $\mathrm{HP}^*_{\mathrm{cont}}(M_k(C^\infty(S^*M)) \otimes \mathbb{C}[\Gamma])$ with $\mathrm{HP}^*_{\mathrm{cont}}(C^\infty(S^*M) \otimes \mathbb{C}[\Gamma])$, and we shall do the same in the proof.

Proposition. 3.3. *The composition*

$$\mathrm{H}^{*-1}(S^*M \times \mathrm{B}\,\Gamma; \mathbb{C}) \hookrightarrow \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(S^*M) \otimes \mathbb{C}[\Gamma]) \xrightarrow{\lambda^*}$$
$$\to \mathrm{HP}^*_{\mathrm{cont}}(C^\infty(S^*M)) \simeq \mathrm{H}^{*-1}(S^*M; \mathbb{C})$$

is $\Phi^{-1} \circ \lambda^* \circ \Phi = (id \times f)^*$.

Proof. Consider as before the morphism $l : \mathcal{G}_{\mathcal{V}} \to S^*M \times R_I$ of groupoids, which defines an injective morphism of algebras $\iota(l) : C^\infty(\mathcal{G}_{\mathcal{V}}) \to C^\infty(S^*M \times R_I) = M_k(C^\infty(S^*M))$, and hence also a morphism

$$\iota(l) \otimes id = \iota(l \times id) : C^\infty(\mathcal{G}_{\mathcal{V}} \times \Gamma) \hookrightarrow C^\infty(S^*M \times R_I \times \Gamma) = M_k(C^\infty(S^*M)) \otimes \mathbb{C}[\Gamma].$$

Then we can write

$$\lambda = \iota(l \times id) \circ \iota(g) \circ \nu,$$

where $g : \mathcal{G}_{\mathcal{V}} \to \mathcal{G}_{\mathcal{V}} \times \Gamma$ is as defined before: $g(x, \alpha, \beta) = (x, \alpha, \beta, \gamma_{\alpha\beta})$.

Because $\nu^* = (t_{Tr})^{-1}$, we have that $\Phi^{-1} \circ \nu^* \circ \Phi = (\mathrm{B}\, t)^{*-1}$, by Theorem 3.1. Also by Theorem 3.1, we have $\iota(g)^* \circ \Phi = \Phi \circ (\mathrm{B}\, g)^*$ and $\iota(l \times id)^* \circ \Phi = \Phi \circ (\mathrm{B}\, l \times id)^*$. This gives then

$$\Phi^{-1} \circ \lambda^* \circ \Phi = \Phi^{-1} \circ \nu^* \circ \Phi \circ (\mathrm{B}\, g)^* \circ (\mathrm{B}\, l \times id)^* = (\mathrm{B}\, t)^{*-1} \circ \Phi \circ (\mathrm{B}\, g)^* \circ (\mathrm{B}\, l \times id)^* = h_0^*.$$

Since Lemma 3.2 states that $h_0 = id \times f$, up to homotopy, the proof is complete. $\square$

3.3. THE ATIYAH–SINGER EXACT SEQUENCE. Let M be a smooth compact manifold (without boundary). We shall denote by $\Psi^k(M)$ the space of classical, order at most k pseudodifferential operators on M. Fix a smooth, nowhere vanishing density on M. Then $\Psi^0(M)$ acts on $L^2(M)$ by bounded operators and, if an operator $T \in \Psi^0(M)$ is compact, then it is of order -1. More precisely, it is known that order -1 pseudodifferential operators satisfy $\Psi^{-1}(M) \subset \mathcal{C}_p = \mathcal{C}_p(L^2(M))$ for any $p > n$. (Recall that $C_p(\mathcal{H})$ is the ideal of p–summable operators on $\mathcal{H}$, equation (8)).

It will be convenient to include all $(n+1)$–summable operators in our calculus, so we let $E_{AS} = \Psi^0(M) + \mathcal{C}_{n+1}$, and obtain in this way an extension of algebras,

$$0 \to \mathcal{C}_{n+1} \to E_{AS} \xrightarrow{\sigma_0} C^\infty(S^*M) \to 0, \tag{40}$$

called *the Atiyah-Singer exact sequence*. The boundary morphisms in periodic cyclic cohomology associated to the Atiyah-Singer exact sequence defines a morphism

$$\partial_{AS} : \mathrm{HP}^*(\mathcal{C}_{n+1}) \to \mathrm{HP}^{*+1}(C^\infty(S^*M)).$$

Let $Tr_n \in \mathrm{HP}^0_{\mathrm{cont}}(\mathcal{C}_{n+1})$ be as in (28) (i.e., $Tr_n(a_0, \ldots, a_{2n}) = CTr(a_0 \ldots a_{2n})$, for some constant C), and denote

$$\mathcal{J}(M) = \partial_{AS}(Tr_n) \in \mathrm{HP}^1_{\mathrm{cont}}(C^\infty(S^*M)) \subset \mathrm{HP}^1(C^\infty(S^*M)), \tag{41}$$

which is justified by Corollary 2.11.

We shall determine $\mathcal{J}(M)$ using Theorem 1.5. In order to do this, we need to make explicit the relation between ch, the Chern character in cyclic homology, and Ch, the classical Chern character as defined, for example, in [27]. Let $E \to M$ be a smooth complex vector bundle, embedded in a trivial bundle: $E \subset M \times \mathbb{C}^N$, and let $e \in M_N(C^\infty(M))$ be the orthogonal projection on E. If we endow E with the connection $ed_{DR}e$, acting on $\Gamma^\infty(E) \subset C^\infty(M)^N$, then the curvature Ω of this connection turns out to be $\Omega = e(d_{DR}e)^2$. The classical Chern character $Ch(E)$ is then the cohomology class of the form $Tr(\exp(\frac{\Omega}{2\pi\imath}))$ in the even (de Rham) cohomology of M. Comparing this definition with the definition of the Chern character in cyclic cohomology via the Hochschild-Kostant-Rosenberg map, we see that the two of them are equal–up to a renormalization with a factor of $2\pi\imath$. (If $\xi \in \mathrm{H}^*(M) = \oplus_k \mathrm{H}^k(M)$ is a cohomology class, we denote by ξ_k its component in $\mathrm{H}^k(M)$.) Explicitly, let $\chi : \mathrm{HP}_i^{\mathrm{cont}}(C_c^\infty(S^*M)) \simeq \oplus_{k\in\mathbb{Z}} \mathrm{H}^{i+2k}(S^*M)$ be the canonical isomorphism induced by the Hochschild-Kostant-Rosenberg map χ, equation (12), then

$$\chi(ch(\xi)) = \sum_{k\in\mathbb{Z}} (2\pi\imath)^m Ch(\xi)_{2m-i} \in \mathrm{H}^{2m-i}(M) \tag{42}$$

for $i \in \{0, 1\}$ and $\xi \in \mathrm{K}_i^{\mathrm{alg}}(C^\infty(M))$. (Note the '$-i$').

Proposition. 3.4. *Let $\mathcal{T}(M) \in \mathrm{H}^{\mathrm{even}}(S^*M)$ be the Todd class of the complexification of T^*M, lifted to S^*M, and $\Phi : \mathrm{H}^{\mathrm{even}}(S^*M) \to \mathrm{HP}^1_{\mathrm{cont}}(C^\infty(S^*M))$ be the isomorphism of Theorem 3.1. Then*

$$\mathcal{J}(M) = (-1)^n \sum_k (2\pi\imath)^{n-k} \Phi(\mathcal{T}(M)_{2k}) \in \mathrm{HP}^1_{\mathrm{cont}}(C^\infty(S^*M)).$$

Proof. We need to verify the equality of two classes in $\mathrm{HP}^1_{\mathrm{cont}}(C^\infty(S^*M))$. It is hence enough to check that their pairings with $ch([u])$ are equal, for any $[u] \in \mathrm{K}_1^{\mathrm{alg}}(C^\infty(S^*M))$, because of the classical result that the Chern character

$$ch : \mathrm{K}_1^{\mathrm{alg}}(C^\infty(S^*M)) \to \mathrm{HP}_1^{\mathrm{cont}}(C^\infty(S^*M))$$

is onto.

If Ind is the index morphism of the Atiyah–Singer exact sequence then the Atiyah-Singer index formula [5] states the equality

$$\mathrm{Ind}[u] = (-1)^n \langle\, Ch[u], \mathcal{T}(M)\,\rangle. \tag{43}$$

Using equation (41) and Theorem 1.5 (see also the discussion following that theorem), we obtain that $\mathrm{Ind}[u] = \langle\, ch[u], \mathcal{J}(M)\,\rangle$. Equations (37) and (43) then complete the proof. □

3.4. The Connes–Moscovici exact sequence and proof of the theorem. We now extend the constructions leading to the Atiyah–Singer exact sequence, equation (40), to covering spaces.

Let M be a smooth compact manifold and let $E_1 = M_k(E) \otimes \mathbb{C}[\Gamma]$, which fits into the exact sequence

$$0 \longrightarrow M_k(\mathcal{C}_{n+1}) \otimes \mathbb{C}[\Gamma] \longrightarrow E_1 \xrightarrow{\sigma_0} M_k(C^\infty(S^*M)) \otimes \mathbb{C}[\Gamma] \longrightarrow 0. \tag{44}$$

Let $\Gamma \to \widetilde{M} \to M$ be a covering of M with Galois group Γ. Using the Mishchenko idempotent p associated to this covering and the injective morphism

$$\lambda : C^\infty(S^*M) \to p(M_k(C^\infty(S^*M)) \otimes \mathbb{C}[\Gamma])p,$$

equation 39, we define the Connes–Moscovici algebra E_{CM} as the fibered product

$$E_{CM} = \{(T,a) \in pE_1p \oplus C^\infty(S^*M), \sigma_0(T) = \lambda(a)\}.$$

By definition, the algebra E_{CM} fits into the exact sequence

$$0 \longrightarrow p\big(M_k(\mathcal{C}_{n+1}) \otimes \mathbb{C}[\Gamma]\big)p \longrightarrow E_{CM} \longrightarrow C^\infty(S^*M) \longrightarrow 0.$$

We now take a closer look at the algebra E_{CM} and the exact sequence it defines. Observe first that p acts on $(L^2(M) \otimes l^2(\Gamma))^k$ and that $p(L^2(M) \otimes l^2(\Gamma))^k \simeq L^2(\widetilde{M})$ via a Γ–invariant isometry. Since E_1 can be regarded as an algebra of operators on $(L^2(M) \otimes l^2(\Gamma))^k$ that commute with the (right) action of Γ, we obtain that E_{CM} can also be interpreted as an algebra of operators commuting with the action of Γ on $L^2(\widetilde{M})$. Using also [11], Lemma 5.1, page 376, this recovers the usual description of E_{CM} that uses properly supported Γ–invariant pseudodifferential operators on $\widetilde{M}$.

Also observe that "M_k" is superfluous in $M_k(\mathcal{C}_{n+1})$ because $M_k(\mathcal{C}_{n+1}) \simeq \mathcal{C}_{n+1}$; actually, even "$p$" is superfluous in $p\big(M_k(\mathcal{C}_{n+1}) \otimes \mathbb{C}[\Gamma]\big)p$ because

$$p\big(M_k(\mathcal{C}_{n+1}) \otimes \mathbb{C}[\Gamma]\big)p \simeq \mathcal{C}_{n+1} \otimes \mathbb{C}[\Gamma]$$

by an isomorphism that is uniquely determined up to an inner automorphism. Thus the Connes–Moscovici extension becomes

$$0 \longrightarrow \mathcal{C}_{n+1} \otimes \mathbb{C}[\Gamma] \longrightarrow E_{CM} \longrightarrow C^\infty(S^*M)) \longrightarrow 0, \tag{45}$$

up to an inner automorphism.

We now proceed as for the Atiyah–Singer exact sequence. The boundary morphisms in periodic cyclic cohomology associated to the Connes–Moscovici extensions defines a map

$$\partial_{CM} : \mathrm{HP}^*(\mathcal{C}_{n+1} \otimes \mathbb{C}[\Gamma]) \to \mathrm{HP}^{*+1}(C^\infty(S^*M)),$$

and the Connes–Moscovici Index Theorem amounts to the identification of the classes

$$\partial_{CM}(Tr_n \otimes \xi) \in \mathrm{HP}^{*+1}_{\mathrm{cont}}(C^\infty(S^*M)) \subset \mathrm{HP}^{*+1}(C^\infty(S^*M)),$$

for cocycles ξ coming from the cohomology of Γ.

In order to determine $\partial_{CM}(Tr_n \otimes \xi)$, we need the following theorem.

THEOREM. 3.5. *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be smooth étale groupoids. Then the diagram*

$$\begin{array}{ccc} \mathrm{H}^{*+n}(\mathrm{B}\,\mathcal{G}_1, \mathfrak{o}_1) \otimes \mathrm{H}^{*+m}(\mathrm{B}\,\mathcal{G}_2, \mathfrak{o}_2) & \xrightarrow{\Phi} & \mathrm{HP}^*_{\mathrm{cont}}(C^\infty_c(\mathcal{G}_2)) \otimes \mathrm{HP}^*_{\mathrm{cont}}(C^\infty_c(\mathcal{G}_2)) \\ {\scriptstyle\times}\big\downarrow & & \big\downarrow{\scriptstyle\otimes} \\ \mathrm{H}^{*+n+m}(\mathrm{B}(\mathcal{G}_1 \times \mathcal{G}_2), \mathfrak{o}) & \xrightarrow{\Phi} & \mathrm{HP}^*_{\mathrm{cont}}(C^\infty_c(\mathcal{G}_1 \times \mathcal{G}_2)) \end{array}$$

is commutative. Here the left product $\times$ is the external product in cohomology and $\mathfrak{o}_1$, $\mathfrak{o}_2$, and $\mathfrak{o}$ are the orientation sheaves.

Proof. The proof is a long but straightforward verification that the sequence of isomorphisms in [7] is compatible with products.

Using [30], Proposition 1.5. (c), page 563, which states that the $\times$-products are compatible with the tensor products of mixed complexes, we replace everywhere cyclic vector spaces by mixed complexes. Then we go through the specific steps of the proof as in [7]. This amounts to verify the following facts:

(i) The Hochschild-Kostant-Rosenberg map χ (equation (12)) transforms the differential $B \otimes 1 + 1 \otimes B$ into the de Rham differential of the product.

(ii) By the Eilenberg-Zilber Theorem [25], the augmentation map ϵ ([7] Proposition 4.2 (1)), and the isomorphism it induces, are compatible with products.

(iii) The chain map f in the Moore isomorphism (see [6], Theorems 4.1 and 4.2, page 32) is compatible with products. This too involves the Eilenberg-Zilber theorem. □

We remark that the proof of the above theorem is easier if both groupoids are of the same "type," i.e., if they are both groups or smooth manifolds, in which case our theorem is part of folklore. However, in the case we shall use this theorem–that of a group and a manifold–there are no significant simplifications: one has to go through all the steps of the proof given above.

LEMMA. 3.6. *Let $\lambda : C^\infty(S^*M) \to M_k(C^\infty(S^*M)) \otimes \mathbb{C}[\Gamma]$ be as defined in* (39) *and $Tr_n \in \mathrm{HP}^0(\mathcal{C}_{n+1})$ be as in* (28)*. Then, for any cyclic cocycle $\eta \in \mathrm{HP}^*_{\mathrm{cont}}(\mathbb{C}[\Gamma])$, we have*

$$\partial_{CM}(Tr_n \otimes \eta) = \lambda^*(\mathcal{J}(M) \otimes \eta) \in \mathrm{HP}^{*+1}_{\mathrm{cont}}(C^\infty(S^*M)) \subset \mathrm{HP}^{*+1}(C^\infty(S^*M)).$$

Proof. Denote by $\partial_1 : \mathrm{HP}^*_{\mathrm{cont}}(\mathcal{C}_{n+1}\otimes\mathbb{C}[\Gamma]) \to \mathrm{HP}^{*+1}(C^\infty(S^*M\otimes\mathbb{C}[\Gamma]))$ the boundary morphism of the exact sequence (44). Using Theorem 2.6, we obtain

$$\partial_1(Tr_n\otimes\eta) = \partial_{AS}(Tr_n)\otimes\eta = \mathcal{J}(M)\otimes\eta \in \mathrm{HP}^{*+1}_{\mathrm{cont}}(C^\infty(S^*M)\otimes\mathbb{C}[\Gamma]) \subset \mathrm{HP}^{*+1}(C^\infty(S^*M)\otimes\mathbb{C}[\Gamma]).$$

Then, the naturality of the boundary map and Theorem 2.10 show that $\partial_{CM} = \lambda^*\circ\partial_1$. This completes the proof. □

Let $\mathcal{T}(M)\in \mathrm{H}^{\mathrm{even}}(S^*M)$ be the Todd class of $TM\otimes\mathbb{C}$ lifted to S^*M and Ch be the classical Chern character on K–Theory, as before. Also, recall that Theorem 3.1 defines an embedding $\Phi : \mathrm{H}^*(\mathrm{B}\,\Gamma) = \mathrm{H}^*(\Gamma) \to \mathrm{HP}^*_{\mathrm{cont}}(\mathbb{C}[\Gamma]) = \mathrm{HP}^*(\mathbb{C}[\Gamma])$.

We are now ready to state Connes–Moscovici's Index Theorem for elliptic systems, see [11][Theorem 5.4], page 379, which computes the "higher index" of a matrix of P of properly supported, order zero, Γ-invariant elliptic pseudodifferential operators on $\widetilde{M}$, with principal symbol the invertible matrix $u=\sigma_0(P)\in M_m(C^\infty(S^*M))$.

Theorem. 3.7 (Connes–Moscovici). *Let $\widetilde{M}\to M$ be a covering with Galois group Γ of a smooth compact manifold M of dimension n, and let $f : S^*M \to \mathrm{B}\,\Gamma$ the continuous map that classifies the covering $S^*\widetilde{M}\to S^*M$. Then, for each cohomology class $\xi\in \mathrm{H}^{2q}(\mathrm{B}\,\Gamma)$ and each $[u]\in K^1(S^*M)$, we have*

$$\tilde{\xi}_*(\mathrm{Ind}[u]) = \frac{(-1)^n}{(2\pi\imath)^q}\langle\, Ch(u)\wedge\mathcal{T}(M)\wedge f^*\xi, [S^*M]\,\rangle,$$

where $\tilde{\xi} = Tr_n\otimes\Phi(\xi)\in \mathrm{HP}^0(\mathcal{C}_{n+1}\otimes\mathbb{C}[\Gamma])$.

Proof. All ingredients of the proof are in place, and we just need to put them together. Let $\xi\in\mathrm{H}^{2q}(\mathrm{B}\,\Gamma)$ and $\tilde{\xi} = Tr_n\otimes\Phi(\xi)$ be as in the statement of the theorem. Then

$$\begin{aligned}
&(-1)^n\tilde{\xi}_*(\mathrm{Ind}[u]) = \\
&= (-1)^n\big(\partial_{CM}\tilde{\xi}\big)_*[u] && \text{by Theorem 1.5}\\
&= (-1)^n\big(\lambda^*(\mathcal{J}(M)\otimes\Phi(\xi))\big)_*[u] && \text{by Lemma 3.6}\\
&= (-1)^n\big(\lambda^*\circ\Phi(\Phi^{-1}(\mathcal{J}(M))\times\xi)\big)_*[u] && \text{by Theorem 3.5}\\
&= (-1)^n\big(\Phi\circ(id\times f)^*(\Phi^{-1}(\mathcal{J}(M))\times\xi)\big)_*[u] && \text{by Proposition 3.3}\\
&= (-1)^n\langle\Phi(\Phi^{-1}(\mathcal{J}(M))\wedge f^*\xi), ch([u])\rangle\\
&= (-1)^n\langle\,\Phi^{-1}(\mathcal{J}(M))\wedge f^*\xi)\wedge\chi(ch[u]), [S^*M]\,\rangle && \text{by equation (37)}\\
&= \sum_{k+j=n-q}(2\pi\imath)^{k-n}\langle\mathcal{T}(M)_{2k}\wedge f^*\xi\wedge\chi(ch[u])_{2j-1}, [S^*M]\rangle && \text{by Proposition 3.4}\\
&= \sum_{k+j=n-q}(2\pi\imath)^{-q}\langle\mathcal{T}(M)_{2k}\wedge f^*\xi\wedge Ch_{2j-1}[u], [S^*M]\rangle && \text{by equation (42)}\\
&= (2\pi\imath)^{-q}\langle Ch[u]\wedge\mathcal{T}(M)\wedge f^*\xi, [S^*M]\rangle.
\end{aligned}$$

The proof is now complete. □

For $q=0$ and $\xi = 1\in\mathrm{H}^0(\mathrm{B}\,\Gamma)\simeq\mathbb{C}$, we obtain that $\tau=\Phi(\xi)$ is the von Neumann trace on $\mathbb{C}[\Gamma]$, that is $\tau(\sum a_\gamma\gamma) = a_e$, the coefficient of the identity, and the above theorem recovers Atiyah's L^2–index theorem for coverings [2]. The reason for obtaining a different constant than in [11] is due to different normalizations. See [19] for a discussion on how to obtain the usual index theorems from the index theorems for elliptic systems.

REFERENCES

[1] M. F. Atiyah. *K-Theory.* W. A. Benjamin, New York, 1967.
[2] M. F. Atiyah. Elliptic operators, discrete subgroups, and von Neumann algebras. *Astérisque*, 32/33:43–72, 1976.
[3] M. F. Atiyah and W. Schmid. A geometric construction of the discrete series. *Invent. Math.*, 42:1–62, 1977.
[4] M. F. Atiyah and I. M. Singer. The index of elliptic operators I. *Ann. of Math.*, 87:484–530, 1968.
[5] M. F. Atiyah and I. M. Singer. The index of elliptic operators III. *Ann. of Math.*, 87:546–604, 1968.
[6] R. Bott. Gelfand-Fuks cohomology and foliations. In *Proceedings of the eleventh annual Symposium at New Mexico State University*, 1973.
[7] J.-L. Brylinski and V. Nistor. Cyclic cohomology of etale groupoids. *K-Theory*, 8:341–365, 1994.
[8] D. Burghelea. The cyclic cohomology of group rings. *Comment. Math. Helvetici*, 60:354–365, 1985.
[9] A. Connes. Cohomology cyclique et foncteurs Ext^n. *C. R. Acad. Sci. Paris, série I*, 296(23):953–958, 1983.
[10] A. Connes. Non-commutative differential geometry. *Publ. Math. IHES*, 62:41–144, 1985.
[11] A. Connes and H. Moscovici. Cyclic cohomology, the Novikov conjecture and hyperbolic groups. *Topology*, 29:345–388, 1990.
[12] J. Cuntz. Bivariante K–Theorie für lokalconvexe Algebren und der Chern-Connes-Charakter. *Doc.Math.J.DMV*, 2:139–182, 1997, `http://www.math.uiuc.edu/documenta/`.
[13] J. Cuntz. Excision in bivariant cyclic theory for topological algebras. To appear.
[14] J. Cuntz and D. Quillen. On excision in periodic cyclic cohomology. *C. R. Acad. Sci. Paris*, 317:917–922, 1993.
[15] J. Cuntz and D. Quillen. On excision in periodic cyclic cohomology II, the general case. *C. R. Acad. Sci. Paris*, 318:11–12, 1994.
[16] J. Cuntz and D. Quillen. Excision in bivariant periodic cyclic cohomology. *Invent. Math.*, 127:67–98, 1997.
[17] A. Haefliger. Groupoïdes d'holonomie et espaces classifiants. *Astérisque*, 116:70–97, 1984.
[18] P. Hilton. Lectures in homological algebra. *Regional Conf. Series in Math.*, 8:1–74, 1971.
[19] J. Kaminker. Pseudodifferential operators and differential structures. *Contemporary Mathematics*, 10:99–128, 1982.
[20] M. Karoubi. Homologie cyclique et K-theorie. *Asterisque*, 149:1–147, 1987.
[21] C. Kassel. Cyclic homology, comodules and mixed complexes. *J. of Algebra*, 107:195–216, 1988.
[22] D. Kazhdan, V. Nistor, and P. Schneider. Hochschild and cyclic homology of finite type algebras. preprint '97.

[23] J.-L Loday. *Cyclic Homology.* Springer, Berlin-Heidelberg-New York, 1992.
[24] J.-L. Loday and D. Quillen. Cyclic homology and the Lie algebra homology of matrices. *Comment. Math. Helvetici*, 59:565–591, 1984.
[25] S. MacLane. *Homology.* Springer, Berlin-Heidelberg-New York, 1963.
[26] J. Milnor. *Algebraic K-Theory*, volume 72 of *Ann. Math. Studies.* Princeton University Press, Princeton, New Jersey, 1974.
[27] J. Milnor and J. Stasheff. *Characteristic Classes*, volume 76 of *Ann. Math. Studies.* Princeton University Press, Princeton, New Jersey, 1974.
[28] V. Nistor. Group cohomology and the cyclic cohomology of crossed products. *Invent. Math.*, 99:411–424, 1990.
[29] V. Nistor. A bivariant Chern character for p-summable quasihomomorphisms. *K-Theory*, 5:193–211, 1991.
[30] V. Nistor. A bivariant Chern-Connes character. *Ann. of Math.*, 138:555–590, 1993.
[31] V. Nistor. Cyclic cohomology of crossed products by algebraic groups. *Invent. Math.*, 112:615–638, 1993.
[32] V. Nistor. On the the Cuntz–Quillen boundary map. *Math. Rep. of Acad. of Sci. Canada*, XVI:203–208, 1995.
[33] V. Nistor. An index theorem for foliated bundles. *J. Funct.Anal.*, 141:421–434, 1996.
[34] D. Quillen. Higher algebraic K-theory I. volume 341 of *Lect. Notes in Math.*, pages 85–174, Berlin-Heidelberg-New York, 1973. Springer.
[35] J. Renault. *A Groupoid approach to C^*-algebras*, volume 793 of *Lect. Notes in Math.* Springer, Berlin-Heidelberg-New York, 1980.
[36] B. Simon. *Trace ideals and their applications*, volume 35 of *London Math. Soc. Lect. Notes.* Cambridge University Press, 1979.
[37] B. L. Tsygan. Homology of matrix Lie algebras over rings and Hochschild homology. *Uspekhi Math. Nauk.*, 38:217–218, 1983.
[38] M. Wodzicki. Excision in cyclic homology and in rational algebraic K-theory. *Ann. of Math.*, 129:591–640, 1989.

Victor Nistor
Pennsylvania State University
Math. Dept.
University Park, PA 16802
USA
nistor@math.psu.edu

On the Group $H^3(F(\psi, D)/F)$

Oleg T. Izhboldin and Nikita A. Karpenko

Received: August 14, 1997

Communicated by Ulf Rehmann

Abstract. Let F be a field of characteristic different from 2, ψ a quadratic F-form of dimension ≥ 5, and D a central simple F-algebra of exponent 2. We denote by $F(\psi, D)$ the function field of the product $X_\psi \times X_D$, where X_ψ is the projective quadric determined by ψ and X_D is the Severi-Brauer variety determined by D. We compute the relative Galois cohomology group $H^3(F(\psi, D)/F, \mathbb{Z}/2\mathbb{Z})$ under the assumption that the index of D goes down when extending the scalars to $F(\psi)$. Using this, we give a new, shorter proof of the theorem [23, Th. 1] originally proved by A. Laghribi, and a new, shorter, and more elementary proof of the assertion [2, Cor. 9.2] originally proved by H. Esnault, B. Kahn, M. Levine, and E. Viehweg.

1991 Mathematics Subject Classification: 19E15, 12G05, 11E81.

Let ψ be a quadratic form and D be an exponent 2 central simple algebra over a field F (always assumed to be of characteristic not 2). Let X_ψ be the projective quadric determined by ψ, X_D the Severi-Brauer variety determined by D, and $F(\psi, D)$ the function field of the product $X_\psi \times X_D$.

A computation of the relative Galois cohomology group

$$H^3(F(\psi, D)/F) \stackrel{\text{def}}{=} \ker\left(H^3(F, \mathbb{Z}/2\mathbb{Z}) \to H^3(F(\psi, D), \mathbb{Z}/2\mathbb{Z})\right)$$

plays a crucial role in obtaining the results of [8] and [10] concerning the problem of isotropy of quadratic forms over the function fields of quadrics.

The group $H^3(F(\psi, D)/F)$ is closely related to the Chow group $\mathrm{CH}^2(X_\psi \times X_D)$ of 2-codimensional cycles on the product $X_\psi \times X_D$. The main result of this paper is the following theorem, where both groups are computed assuming $\dim \psi \geq 5$ and the index of D goes down when extending the scalars to the function field of ψ:

Theorem 0.1. *Let D be a central simple F-algebra of exponent 2. Let ψ be a quadratic form of dimension ≥ 5. Suppose that* $\operatorname{ind} D_{F(\psi)} < \operatorname{ind} D$. *Then* $\operatorname{Tors}\mathrm{CH}^2(X_\psi \times X_D) = 0$ *and* $H^3(F(\psi, D)/F) = [D] \cup H^1(F)$.

A proof is given in §8. The essential part of the proof is Theorem 6.9, dealing with the special case where D is a division algebra of degree 8. This theorem has two applications in the theory of quadratic forms. The first one is a new, shorter proof of the following assertion, originally proved by A. Laghribi ([23, Th. 1]):

Corollary 0.2. *Let* $\phi \in I^2(F)$ *be an 8-dimensional quadratic form such that* $\operatorname{ind} C(\phi) = 8$. *Let* ψ *be a quadratic form of dimension* ≥ 5 *such that* $\phi_{F(\psi)}$ *is isotropic. Then there exists a half-neighbor* ϕ^* *of* ϕ *such that* $\psi \subset \phi^*$.

The other application we demonstrate is a new, shorter, and more elementary proof of the assertion, originally proved by H. Esnault, B. Kahn, M. Levine, and E. Viehweg ([2, Cor. 9.2]):

Corollary 0.3. *Let* $\phi \in I^2(F)$ *be any quadratic form such that* $\operatorname{ind} C(\phi) \geq 8$. *Let* A *be a central simple* F*-algebra Brauer equivalent to* $C(\phi)$ *and let* $F(A)$ *be the function field of the Severi-Brauer variety of* A. *Then* $\phi_{F(A)} \notin I^4(F(A))$. *In particular,* $\phi_{F(A)}$ *is not hyperbolic. Moreover, if* $\dim \phi = 8$ *then* $\phi_{F(A)}$ *is anisotropic.*

Our proofs of Corollaries 0.2 and 0.3 are given in §7.

An important part in the proof of Theorem 6.9 is played by the formula of Proposition 4.5, which is in fact applicable to a wide class of algebraic varieties.

A computation of the group $H^3(F(\psi, D)/F)$ in some cases not covered by Theorem 0.1 is given in [8] and [10].

1. Terminology, notation, and backgrounds

1.1. Quadratic forms. Mainly, we use notation of [24] and [30]. However there is a slight difference: we denote by $\langle\langle a_1, \dots, a_n\rangle\rangle$ the n-fold Pfister form

$$\langle 1, -a_1\rangle \otimes \cdots \otimes \langle 1, -a_n\rangle .$$

The set of all n-fold Pfister forms over F is denoted by $P_n(F)$; $GP_n(F)$ is the set of forms similar to a form from $P_n(F)$.

We recall that a quadratic form ψ is called a *(Pfister) neighbor* (of a Pfister form π), if it is similar to a subform in π and $\dim \phi > \frac{1}{2} \dim \pi$. Two quadratic forms ϕ and ϕ^* are *half-neighbors*, if $\dim \phi = \dim \phi^*$ and there exists $s \in F^*$ such that the sum $\phi \perp s\phi^*$ is similar to a Pfister form.

For a quadratic form ϕ of dimension ≥ 3, we denote by X_ϕ the projective variety given by the equation $\phi = 0$ and we set $F(\phi) = F(X_\phi)$.

1.2. Generic splitting tower. Let γ be a non-hyperbolic quadratic form over F. Put $F_0 \stackrel{\text{def}}{=} F$ and $\gamma_0 \stackrel{\text{def}}{=} \gamma_{an}$. For $i \geq 1$ let $F_i \stackrel{\text{def}}{=} F_{i-1}(\gamma_{i-1})$ and $\gamma_i \stackrel{\text{def}}{=} ((\gamma_{i-1})_{F_i})_{an}$. The smallest h such that $\dim \gamma_h \leq 1$ is called the *height* of γ. The sequence $F_0, F_1, \dots, F_h$ is called the *generic splitting tower* of γ ([21]). We need some properties of the fields F_s:

Lemma 1.3 ([22]). *Let* M/F *be a field extension such that* $\dim(\gamma_M)_{an} = \dim \gamma_s$. *Then the field extension* MF_s/M *is purely transcendental.*

The following proposition is a consequence of the index reduction formula [25].

Proposition 1.4 (see [6, Th. 1.6] or [5, Prop. 2.1]). *Let* $\phi \in I^2(F)$ *be a quadratic form with* $\operatorname{ind}(C(\phi)) \geq 2^r > 1$. *Then there is* s $(0 \leq s \leq h(\phi))$ *such that* $\dim \phi_s = 2r + 2$ *and* $\operatorname{ind} C(\phi_s) = 2^r$.

Corollary 1.5. *Let* $\phi \in I^2(F)$ *be a quadratic form with* $\operatorname{ind}(C(\phi)) \geq 8$. *Then there is* s $(0 \leq s \leq h(\phi))$ *such that* $\dim \phi_s = 8$ *and* $\operatorname{ind} C(\phi_s) = 8$. □

1.6. Central simple algebras. We are working with finite-dimensional associative algebras over a field. Let D be a central simple F-algebra. We denote by X_D the Severi-Brauer variety of D and by $F(D)$ the function field $F(X_D)$.

For another central simple F-algebra D' and for a quadratic F-form ψ of dimension ≥ 3, we set $F(D', D) \overset{\text{def}}{=} F(X_{D'} \times X_D)$ and $F(\psi, D) \overset{\text{def}}{=} F(X_\psi \times X_D)$.

1.7. Galois cohomology. By $H^*(F)$ we denote the graded ring of Galois cohomology

$$H^*(F, \mathbb{Z}/2\mathbb{Z}) = H^*(\mathrm{Gal}(F_{\mathrm{sep}}/F), \mathbb{Z}/2\mathbb{Z}).$$

For any field extension L/F, we set $H^*(L/F) \overset{\text{def}}{=} \ker(H^*(F) \to H^*(L))$.

We use the standard canonical isomorphisms $H^0(F) = \mathbb{Z}/2\mathbb{Z}$, $H^1(F) = F^*/F^{*2}$, and $H^2(F) = \mathrm{Br}_2(F)$.

We also work with the cohomology groups $H^n(F, \mathbb{Q}/\mathbb{Z}(i))$, $i = 0, 1, 2$ (see e.g. [12] for the definition). For any field extension L/F, we set

$$H^*(L/F, \mathbb{Q}/\mathbb{Z}(i)) \overset{\text{def}}{=} \ker\left(H^*(F, \mathbb{Q}/\mathbb{Z}(i)) \to H^*(L, \mathbb{Q}/\mathbb{Z}(i))\right).$$

For $n = 1, 2, 3$, the group $H^n(F)$ is naturally identified with

$$\mathrm{Tors}_2\, H^n(F, \mathbb{Q}/\mathbb{Z}(n-1)).$$

1.8. K-theory and Chow groups. We are mainly working with smooth algebraic varieties over a field, although the smoothness assumption is not always essential.

Let X be a smooth algebraic F-variety. The Grothendieck ring of X is denoted by $K(X)$. This ring is supplied with the filtration "by codimension of support" (which respects multiplication); the adjoint graded ring is denoted by $G^*K(X)$. There is a canonical surjective homomorphism of the graded Chow ring $\mathrm{CH}^*(X)$ onto $G^*K(X)$; its kernel consists only of torsion elements and is trivial in the 0-th, 1-st and 2-nd graded components ([32, §9]). In particular we have the following

Lemma 1.9. *The homomorphism* $\mathrm{CH}^i(X) \to G^iK(X)$ *is bijective if at least one of the following conditions holds:*

- $i = 0$, 1, *or* 2,
- $\mathrm{CH}^i(X)$ *is torsion-free.*

Let X be a variety over F and E/F be a field extension. We denote by $i_{E/F}$ the restriction homomorphism $K(X) \to K(X_E)$. We use the same notation for the restriction homomorphisms $\mathrm{CH}^*(X) \to \mathrm{CH}^*(X_E)$ and $G^*K(X) \to G^*K(X_E)$. Note that for any projective homogeneous variety X, the homomorphism $i_{E/F}: K(X) \to K(X_E)$ is injective by [27].

1.10. Other notations. We denote by $\bar{F}$ a separable closure of the field F. The order of a set S is denoted by $|S|$ (if S is infinite, we set $|S| \overset{\text{def}}{=} \infty$).

2. The group $\operatorname{Tors} G^*K(X)$

Lemma 2.1. *Let X be a variety over F and E/F be a field extension such that the homomorphism $i_{E/F} : K(X) \to K(X_E)$ is injective and the factor group $K(X_E)/i_{E/F}(K(X))$ is finite. Then*

$$|\ker(G^*K(X) \to G^*K(X_E)| = \frac{|G^*K(X_E)/i_{E/F}(G^*K(X))|}{|K(X_E)/i_{E/F}(K(X))|}$$

Proof. The proof is the same as the proof of [15, Prop. 2]. □

Lemma 2.2. *Let X be a variety, i be an integer, and E/F be a field extension such that the group $G^iK(X_E)$ is torsion-free. Then*

$$\ker(G^iK(X) \to G^iK(X_E)) = \operatorname{Tors} G^iK(X) .$$

Proof. Since $G^iK(X_E)$ is torsion-free, one has $\ker(G^iK(X) \to G^iK(X_E)) \supset \operatorname{Tors} G^iK(X)$.

To prove the inverse inclusion, let us take an intermediate field E_0 such that the extension E_0/F is purely transcendental while the extension E/E_0 is algebraic. The specialization argument shows that the homomorphism $G^iK(X) \to G^iK(X_{E_0})$ is injective; the transfer argument shows that $\ker(G^iK(X_{E_0}) \to G^iK(X_E)) \subset \operatorname{Tors} G^iK(X_{E_0})$. Therefore $\ker(G^iK(X) \to G^iK(X_E)) \subset \operatorname{Tors} G^iK(X)$. □

Lemma 2.3. *Let X be a smooth variety, i be an integer, and E/F be a field extension such that the group $\mathrm{CH}^i(X_E)$ is torsion-free. Then*

- $\mathrm{CH}^i(X_E) \simeq G^iK(X_E)$ *(and hence the group $G^iK(X_E)$ is torsion-free),*
- $\mathrm{CH}^i(X_E)/i_{E/F}(\mathrm{CH}^i(X)) \simeq G^iK(X_E)/i_{E/F}(G^iK(X))$.

Proof. The first assertion is contained in Lemma 1.9. The canonical homomorphism $\mathrm{CH}^i(X_E) \to G^iK(X_E)$ induces a homomorphism

$$\mathrm{CH}^i(X_E)/i_{E/F}(\mathrm{CH}^i(X)) \to G^iK(X_E)/i_{E/F}(G^iK(X))$$

which is bijective since $\mathrm{CH}^i(X_E) \to G^iK(X_E)$ is bijective and $\mathrm{CH}^i(X) \to G^iK(X)$ is surjective. □

Proposition 2.4. *Suppose that a smooth F-variety X and a field extension E/F satisfy the following three conditions:*

- *the homomorphism $i_{E/F} : K(X) \to K(X_E)$ is injective,*
- *the factor group $K(X_E)/i_{E/F}(K(X))$ is finite,*
- *the group $\mathrm{CH}^*(X_E)$ is torsion-free.*

Then

$$|\operatorname{Tors} G^*K(X)| = \frac{|G^*K(X_E)/i_{E/F}(G^*K(X))|}{|K(X_E)/i_{E/F}(K(X))|} = \frac{|\mathrm{CH}^*(X_E)/i_{E/F}(\mathrm{CH}^*K(X))|}{|K(X_E)/i_{E/F}(K(X))|}$$

Proof. It is an obvious consequence of Lemmas 2.1, 2.2, and 2.3. □

3. Auxiliary lemmas

For an Abelian group A we use the notation $\operatorname{rk}(A) = \dim_{\mathbb{Q}}(A \otimes_{\mathbb{Z}} \mathbb{Q})$.

LEMMA 3.1. *Let $A_0 \subset A$, $B_0 \subset B$ be free Abelian groups such that $\operatorname{rk} A_0 = \operatorname{rk} A = r_A$, $\operatorname{rk} B_0 = \operatorname{rk} B = r_B$. Then*

$$\left|\frac{A \otimes_{\mathbb{Z}} B}{A_0 \otimes_{\mathbb{Z}} B_0}\right| = \left|\frac{A}{A_0}\right|^{r_B} \cdot \left|\frac{B}{B_0}\right|^{r_A}.$$

Proof. One has

$$(A \otimes B)/(A_0 \otimes B) \simeq (A/A_0) \otimes B \simeq (A/A_0) \otimes \mathbb{Z}^{r_B} \simeq (A/A_0)^{r_B},$$
$$(A_0 \otimes B)/(A_0 \otimes B_0) \simeq A_0 \otimes (B/B_0) \simeq \mathbb{Z}^{r_A} \otimes (B/B_0) \simeq (B/B_0)^{r_A}.$$

Therefore,

$$\left|\frac{A \otimes B}{A_0 \otimes B_0}\right| = \left|\frac{A \otimes B}{A_0 \otimes B}\right| \cdot \left|\frac{A_0 \otimes B}{A_0 \otimes B_0}\right| = \left|\frac{A}{A_0}\right|^{r_B} \cdot \left|\frac{B}{B_0}\right|^{r_A}.$$

□

The following lemma is well-known.

LEMMA 3.2. *Let A be an Abelian group with a finite filtration $A = \mathcal{F}^0 A \supset \mathcal{F}^1 A \supset \cdots \supset \mathcal{F}^k A = 0$. Let B be a subgroup of A with the filtration $\mathcal{F}^p B = B \cap \mathcal{F}^p A$. Let $G^*A = \bigoplus_{p \geq 0} \mathcal{F}^p A/\mathcal{F}^{p+1} A$ and $G^*B = \bigoplus_{p \geq 0} \mathcal{F}^p B/\mathcal{F}^{p+1} B$. Then*

- *$|A/B| = |G^*A/G^*B|$,*
- *if A is a finitely generated group then $\operatorname{rk} G^*A = \operatorname{rk} A$.* □

In the following lemma the term "ring" means a *commutative ring with unit.*

LEMMA 3.3. *Let A and B be rings whose additive groups are finitely generated Abelian groups. Let I be a nilpotent ideal of A such that $A/I \simeq \mathbb{Z}$. Let R be a subring of $A \otimes_{\mathbb{Z}} B$ and A_R be a subring of A such that $A_R \otimes 1 \subset R$. Then the following inequality holds*

$$\left|\frac{A \otimes_{\mathbb{Z}} B}{R}\right| \leq \left|\frac{A}{A_R}\right|^{r_B} \cdot \left|\frac{A \otimes_{\mathbb{Z}} B}{R + (I \otimes_{\mathbb{Z}} B)}\right|^{r_A}$$

where $r_A = \operatorname{rk} A$ and $r_B = \operatorname{rk} B$.

Proof. Let us denote by B_R the image of R under the following composition $A \otimes B \to (A/I) \otimes B \simeq \mathbb{Z} \otimes B \simeq B$. Obviously,

$$\left|\frac{A \otimes_{\mathbb{Z}} B}{R + (I \otimes_{\mathbb{Z}} B)}\right| = \left|\frac{B}{B_R}\right|.$$

For any $p \geq 0$ we set $\mathcal{F}^p A = \{a \in A \,|\, \exists m \in \mathbb{N} \text{ such that } ma \in I^p\}$. Clearly, $\operatorname{Tors}(A/\mathcal{F}^p A) = 0$, and so $A/\mathcal{F}^p$ is a free Abelian group. Therefore all factor groups $\mathcal{F}^p A/\mathcal{F}^{p+1} A$ $(p = 0, 1, \dots)$ are free Abelian. Since $A/I \simeq \mathbb{Z}$, it follows that $\mathcal{F}^1 A = I$. Thus $A/\mathcal{F}^1 A \simeq \mathbb{Z}$. Since I is a nilpotent ideal of A, there exists k such that $I^k = 0$. Then $\mathcal{F}^k A = 0$. Thus the filtration $A = \mathcal{F}^0 A \supset \mathcal{F}^1 A \supset \mathcal{F}^2 A \supset \dots$ is finite and results of Lemma 3.2 can be applied.

Let $\mathcal{F}^p A_R \stackrel{\text{def}}{=} R \cap \mathcal{F}^p A$, $\mathcal{F}^p(A \otimes B) \stackrel{\text{def}}{=} \operatorname{im}(\mathcal{F}^p A \otimes B \to A \otimes B)$, and $\mathcal{F}^p R \stackrel{\text{def}}{=} R \cap \mathcal{F}^p(A \otimes B)$. If K is one of the rings A, A_R, $A \otimes B$, or R, we set $G^p K \stackrel{\text{def}}{=} \mathcal{F}^p K/\mathcal{F}^{p+1} K$ and $G^*K \stackrel{\text{def}}{=} \bigoplus_{p \geq 0} \mathcal{F}^p K/\mathcal{F}^{p+1} K$. Obviously, $\mathcal{F}^p K \cdot \mathcal{F}^q K \subset \mathcal{F}^{p+q} K$ for all p and q.

Therefore, $K = \mathcal{F}^0 K \supset \mathcal{F}^1 K \supset \cdots \supset \mathcal{F}^p K \supset \dots$ is a ring filtration. Hence, the adjoint graded group G^*K has a graded ring structure. Since the additive group of B is free, we have a natural ring isomorphism $G^*A \otimes B \simeq G^*(A \otimes B)$.

Since $A_R \otimes 1 \subset R$, we have $G^*A_R \otimes 1 \subset G^*R$. Clearly $G^0(A \otimes B) = (A/I) \otimes B$, and G^0R coincides with the image of the composition $R \to A \otimes B \to (A/I) \otimes B$. By definition of B_R, one has $G^0R = 1_{G^*A} \otimes B_R$ (here 1_{G^*A} denotes the unit of the ring G^*A). Therefore $1_{G^*A} \otimes B_R \subset G^*R$. Since $G^*A_R \otimes 1 \subset G^*R$, $1_{G^*A} \otimes B_R \subset G^*R$, and G^*R is a subring of $G^*A \otimes B$, we have $G^*A_R \otimes B_R \subset G^*R$. Therefore $|G^*(A \otimes B)/G^*R| \leq |(G^*A \otimes B)/(G^*A_R \otimes B_R)|$. Applying Lemmas 3.1 and 3.2, we have

$$\left|\frac{A \otimes B}{R}\right| = \left|\frac{G^*(A \otimes B)}{G^*R}\right| \leq \left|\frac{G^*A \otimes B}{G^*A_R \otimes B_R}\right| = \left|\frac{G^*A}{G^*A_R}\right|^{r_B} \cdot \left|\frac{B}{B_R}\right|^{r_A} =$$
$$= \left|\frac{A}{A_R}\right|^{r_B} \cdot \left|\frac{B}{B_R}\right|^{r_A} = \left|\frac{A}{A_R}\right|^{r_B} \cdot \left|\frac{A \otimes_{\mathbb{Z}} B}{R + (I \otimes_{\mathbb{Z}} B)}\right|^{r_A}.$$

□

4. On the group $\mathrm{CH}^*(X \times Y)$

Let X be a smooth variety. We denote by $\mathcal{F}^p\mathrm{CH}^*(X)$ the group

$$\bigoplus_{i \geq p} \mathrm{CH}^i(X)\ .$$

Let Y be another smooth variety. For a subgroup A of $\mathrm{CH}^*(X)$ and a subgroup B of $\mathrm{CH}^*(Y)$, we denote by $A \boxtimes B$ the image of the composition $A \otimes B \to \mathrm{CH}^*(X) \otimes \mathrm{CH}^*(Y) \to \mathrm{CH}^*(X \times Y)$.

The following assertion is evident (see also [20, §3] or [11]).

Proposition 4.1. *Let X and Y be smooth varieties over F. Then*

- *the natural homomorphism $\mathrm{CH}^*(X \times Y) \to \mathrm{CH}^*(Y_{F(X)})$ is surjective,*
- *the kernel of the homomorphism $\mathrm{CH}^*(X \times Y) \to \mathrm{CH}^*(Y_{F(X)})$ contains the group $\mathcal{F}^1\mathrm{CH}^*(X) \boxtimes \mathrm{CH}^*(Y)$.*

□

Corollary 4.2. *If the natural homomorphism $\mathrm{CH}^*(X) \otimes \mathrm{CH}^*(Y) \to \mathrm{CH}^*(X \times Y)$ is bijective and $\mathrm{CH}^*(Y)$ is torsion-free, then the homomorphism $\mathrm{CH}^*(X \times Y) \to \mathrm{CH}^*(Y_{F(X)})$ induces an isomorphism*

$$\frac{\mathrm{CH}^*(X \times Y)}{\mathcal{F}^1\mathrm{CH}^*(X) \boxtimes \mathrm{CH}^*(Y)} \to \mathrm{CH}^*(Y_{F(X)}).$$

Proof. Since $\mathrm{CH}^*(X) \otimes \mathrm{CH}^*(Y) \simeq \mathrm{CH}^*(X \times Y)$ and $\mathrm{CH}^*(X)/\mathcal{F}^1\mathrm{CH}^*(X) \simeq \mathrm{CH}^0(X)$, the factor group $\mathrm{CH}^*(X \times Y)/(\mathcal{F}^1\mathrm{CH}^*(X) \boxtimes \mathrm{CH}^*(Y))$ is isomorphic to $\mathrm{CH}^0(X) \otimes_{\mathbb{Z}} \mathrm{CH}^*(Y) \simeq \mathbb{Z} \otimes_{\mathbb{Z}} \mathrm{CH}^*(Y) \simeq \mathrm{CH}^*(Y)$. Thus, it is sufficient to prove that the homomorphism $\mathrm{CH}^*(Y) \to \mathrm{CH}^*(Y_{F(X)})$ is injective. This is obvious since $\mathrm{CH}^*(Y)$ is torsion-free. □

Corollary 4.3. *Let X and Y be smooth varieties and E/F be a field extension such that the natural homomorphism $\mathrm{CH}^*(X_E) \otimes \mathrm{CH}^*(Y_E) \to \mathrm{CH}^*(X_E \times Y_E)$ is bijective and $\mathrm{CH}^*(Y_E)$ is torsion-free. Then there exists an isomorphism*

$$\frac{\mathrm{CH}^*(X_E \times Y_E)}{i_{E/F}(\mathrm{CH}^*(X \times Y)) + \mathcal{F}^1\mathrm{CH}^*(X_E) \boxtimes \mathrm{CH}^*(Y_E)} \simeq \frac{\mathrm{CH}^*(Y_{E(X)})}{i_{E(X)/F(X)}(\mathrm{CH}^*(Y_{F(X)}))}$$

Proof. Obvious in view of Corollary 4.2. □

Remark 4.4. It was noticed by the referee that the conditions of Corollary 4.3 (which appear also in Proposition 4.5) hold, if the variety Y_E possess a cellular decomposition (see e.g. [13, Def. 3.2] for the definition of cellular decomposition). In the case of complete varieties X and Y, this statement follows e.g. from [19, Th. 6.5]. In the present paper, we shall apply Corollary 4.3 only to the case where Y_E is isomorphic to a projective space.

Proposition 4.5. *Let X and Y be smooth varieties over F and E/F be a field extension such that the following conditions hold*

- $\mathrm{CH}^*(X_E)$ *is a free Abelian group of rank r_X,*
- $\mathrm{CH}^*(Y_E)$ *is a free Abelian group of rank r_Y,*
- *the canonical homomorphism $\mathrm{CH}^*(X_E) \otimes_{\mathbb{Z}} \mathrm{CH}^*(Y_E) \to \mathrm{CH}^*(X_E \times Y_E)$ is an isomorphism.*

Then

$$\left|\frac{\mathrm{CH}^*(X_E \times Y_E)}{i_{E/F}(\mathrm{CH}^*(X \times Y))}\right| \le \left|\frac{\mathrm{CH}^*(X_E)}{i_{E/F}(\mathrm{CH}^*(X))}\right|^{r_Y} \cdot \left|\frac{\mathrm{CH}^*(Y_{E(X)})}{i_{E(X)/F(X)}(\mathrm{CH}^*(Y_{F(X)}))}\right|^{r_X}.$$

Proof. Let $A = \mathrm{CH}^*(X_E)$, $A_R = i_{E/F}(\mathrm{CH}^*(X))$ and $I = \bigoplus_{p>0} \mathrm{CH}^p(X_E) = \mathcal{F}^1\mathrm{CH}^*(X_E)$. Let $B = \mathrm{CH}^*(Y_E)$. By our assumption, we have $\mathrm{CH}^*(X_E \times Y_E) \simeq A \otimes_{\mathbb{Z}} B$. We denote by R the image of the composition $\mathrm{CH}^*(X \times Y) \to \mathrm{CH}^*(X_E \otimes Y_E) \simeq A \otimes_{\mathbb{Z}} B$. Clearly, all conditions of Lemma 3.3 hold. Moreover,

$$\left|\frac{\mathrm{CH}^*(X_E \times Y_E)}{i_{E/F}(\mathrm{CH}^*(X \times Y))}\right| = \left|\frac{A \otimes_{\mathbb{Z}} B}{R}\right| \quad \text{and} \quad \left|\frac{\mathrm{CH}^*(X_E)}{i_{E/F}(\mathrm{CH}^*(X))}\right| = \left|\frac{A}{A_R}\right|.$$

By Corollary 4.3 we have

$$\left|\frac{A \otimes_{\mathbb{Z}} B}{R + (I \otimes_{\mathbb{Z}} B)}\right| = \left|\frac{\mathrm{CH}^*(Y_{E(X)})}{i_{E(X)/F(X)}(\mathrm{CH}^*(Y_{F(X)}))}\right|.$$

To complete the prove it suffices to apply Lemma 3.3. □

5. The group $\operatorname{Tors}\mathrm{CH}^2(X_\psi \times X_D)$

The aim of this section is Corollary 5.6.

Proposition 5.1 (see [14, §2.1]). *Let ψ be a $(2n+1)$-dimensional quadratic form over a separably closed field. Set $X \overset{\text{def}}{=} X_\psi$ and $d \overset{\text{def}}{=} \dim X = 2n-1$. Then for all $0 \le p \le d$ the group $\mathrm{CH}^p(X)$ is canonically isomorphic to $\mathbb{Z}$ (for other p the group $\mathrm{CH}^p(X)$ is trivial). Moreover,*

- *if $0 \le p < n$, then $\mathrm{CH}^p(X) = \mathbb{Z} \cdot h^p$, where $h \in \mathrm{CH}^1(X)$ denotes the class of a hyperplane section of X;*

- *if* $n \le p \le d$, *then* $\mathrm{CH}^p(X) = \mathbb{Z} \cdot l_{d-p}$, *where* l_{d-p} *denotes the class of a linear subspace in* X *of dimension* $d-p$, *besides* $2l_{d-p} = h^p$.

Corollary 5.2. *Let* ψ *be a* $(2n+1)$*-dimensional quadratic form over* F *and let* $X = X_\psi$. *Then*

- $\mathrm{CH}^*(X_{\bar F})$ *is a free Abelian group of rank* $2n$,
- *if* $0 \le p < n$ *then* $|\mathrm{CH}^p(X_{\bar F})/i_{\bar F/F}(\mathrm{CH}^p(X))| = 1$,
- *if* $n \le p \le 2n-1$ *then* $|\mathrm{CH}^p(X_{\bar F})/i_{\bar F/F}(\mathrm{CH}^p(X))| \le 2$,
- $|\mathrm{CH}^*(X_{\bar F})/i_{\bar F/F}(\mathrm{CH}^*(X))| \le 2^n$. □

Proposition 5.3. *Let* D *be a central simple* F*-algebra of exponent* 2 *and of degree* 8. *Let* $E/L/F$ *be field extensions such that* $\operatorname{ind} D_L = 4$ *and* $\operatorname{ind} D_E = 1$. *Let* $Y = \mathrm{SB}(D)$. *For any* $0 \le p \le \dim Y = 7$, *the group* $\mathrm{CH}^p(Y_E)$ *is canonically isomorphic to* $\mathbb{Z}$. *Moreover, the image of the homomorphism* $i_{E/L} : \mathrm{CH}^p(Y_L) \to \mathrm{CH}^p(Y_E) \simeq \mathbb{Z}$ *contains* 1 *if* $p = 0, 4$; 2 *if* $p = 1, 2, 5, 6$; 4 *if* $p = 3, 7$.

Proof. Since $\deg D = 8$ and $\operatorname{ind} D_E = 1$, Y_E is isomorphic to $\mathbb{P}^7_E$. Hence, the group $\mathrm{CH}^p(Y_E) \cong \mathrm{CH}^p(\mathbb{P}^7_E)$ (where $p = 0, \dots, 7$) is generated by the class h^p of a linear subspace ([4]).

The rest part of the proposition is contained in [16, Th.]. For the reader's convenience, we also give a direct construction of the elements required. The class of Y_L itself gives $1 \in i_{E/L}(\mathrm{CH}^0(Y_L))$. Let ξ be the tautological line bundle on the projective space $\mathbb{P}^7_E \simeq Y_E$. Since $\exp D = 2$, the bundle $\xi^{\otimes 2}$ is defined over F and, in particular, over L. Its first Chern class gives $2 \in i_{E/L}(\mathrm{CH}^1(Y_L))$. Since $\operatorname{ind} D_L = 4$, the bundle $\xi^{\oplus 4}$ is defined over L. Its second Chern class gives $6 \in i_{E/L}(\mathrm{CH}^2(Y_L))$.[1] Thus $2 \in i_{E/L}(\mathrm{CH}^2(Y_L))$. The third Chern class of $\xi^{\oplus 4}$ gives $4 \in i_{E/L}(\mathrm{CH}^3(Y_L))$. The fourth Chern class of $\xi^{\oplus 4}$ gives $1 \in i_{E/L}(\mathrm{CH}^4(Y_L))$. Finally, taking the product of the cycles constructed in codimensions 1, 2, and 3 with the cycle of codimension 4, one gets the cycles of codimensions 5, 6, and 7 required. □

Corollary 5.4. *Under the condition of Proposition 5.3, we have*

$$|\mathrm{CH}^*(Y_E)/i_{E/L}(\mathrm{CH}^*(Y_L))| \le 256 .$$

Proof. $\prod_{p=0}^{7} |\mathrm{CH}^p(Y_E)/i_{E/L}(\mathrm{CH}^p(Y_L))| \le 1 \cdot 2 \cdot 2 \cdot 4 \cdot 1 \cdot 2 \cdot 2 \cdot 4 = 256$. □

Proposition 5.5. *Let* D *be a central division* F*-algebra of degree* 8 *and exponent* 2. *Let* ψ *be a* 5*-dimensional quadratic* F*-form. Suppose that* $D_{F(\psi)}$ *is not a skewfield. Then* $\operatorname{Tors} G^*K(X_\psi \times X_D) = 0$.

Proof. Let $X = X_\psi$ and $Y = X_D$. Corollary 5.2 shows that $\mathrm{CH}^*(X_{\bar F})$ is a free abelian group of rank $r_X = 4$ and $|\mathrm{CH}^*(X_{\bar F})/i_{\bar F/F}(\mathrm{CH}^*(X))| \le 2^2 = 4$.

Since D is a division algebra of degree 8 and $D_{F(\psi)}$ is not division algebra, it follows that $\operatorname{ind} D_{F(X)} = 4$. Applying Corollary 5.4 to the case $L = F(X)$, $E = \bar F(X)$, we have $|\mathrm{CH}^*(Y_{\bar F(X)})/i_{\bar F(X)/F(X)}(\mathrm{CH}^*(Y_{F(X)}))| \le 256$.

[1] In fact, it is enough only to know that the *Grothendieck classes* of the bundles $\xi^{\otimes 2}$ and $\xi^{\oplus 4}$ are in $K(Y_L)$ what can be also seen from the computation of the K-theory.

Since $Y_{\bar F} = \mathrm{SB}(D_{\bar F}) \simeq \mathbb{P}^7_{\bar F}$, the group $\mathrm{CH}^*(Y_{\bar F})$ is a free Abelian of rank $r_Y = 8$ and $\mathrm{CH}^*(X_{\bar F}) \otimes \mathrm{CH}^*(Y_{\bar F}) \simeq \mathrm{CH}^*(X_{\bar F} \times Y_{\bar F})$ (see [3, Prop. 14.6.5]). Thus all conditions of Proposition 4.5 hold for X, Y, $E = \bar F$ and we have

$$\left| \frac{\mathrm{CH}^*(X_{\bar F} \times Y_{\bar F})}{i_{\bar F/F}(\mathrm{CH}^*(X \times Y))} \right| \le 4^8 \cdot 256^4 = 2^{48}.$$

Using [29, Th. 4.1 of §8] and [33, Th. 9.1], we get a natural (with respect to extensions of F) isomorphism

$$K(X \times Y) \simeq K\big((F^{\times 3} \times C) \otimes_F (F^{\times 4} \times D^{\times 4})\big) \simeq$$
$$\simeq K\big(F^{\times 12} \times C^{\times 4} \times D^{\times 12} \times (C \otimes_F D)^{\times 4}\big)$$

where $C \overset{\text{def}}{=} C_0(\psi)$ is the even Clifford algebra of ψ. Note that C is a central simple F-algebra of the degree 2^2. Since $D_{F(\psi)}$ is not a skew field, [25, Th. 1] states that $D \simeq C \otimes_F B$ with some central division F-algebra B. Therefore, $\operatorname{ind} C = \deg C = 2^2$ and $\operatorname{ind} C \otimes D = \operatorname{ind} B = \deg B = 2$. Hence

$$\left| \frac{K(X_{\bar F} \times Y_{\bar F})}{i_{\bar F/F}(K(X \times Y))} \right| = (\operatorname{ind} C)^4 \cdot (\operatorname{ind} D)^{12} \cdot (\operatorname{ind} C \otimes D)^4 = 2^{2\cdot 4 + 3 \cdot 12 + 1 \cdot 4} = 2^{48}\ .$$

Applying Proposition 2.4 to the variety $X \times Y$ and $E = \bar F$, we have

$$|\operatorname{Tors} G^* K(X \times Y)| = \frac{|\mathrm{CH}^*(X_{\bar F} \times Y_{\bar F})/i_{\bar F/F}(\mathrm{CH}^*(X \times Y))|}{|K(X_{\bar F} \times Y_{\bar F})/i_{\bar F/F}(K(X \times Y))|} \le \frac{2^{48}}{2^{48}} = 1\ .$$

Therefore, $\operatorname{Tors} G^* K(X \times Y) = 0$. ☐

Applying Lemma 1.9 we get the following

Corollary 5.6. *Under the condition of Proposition 5.5, the group* $\mathrm{CH}^2(X_\psi \times X_D)$ *is torsion-free.* ☐

6. A special case of Theorem 0.1

In this section we prove Theorem 0.1 in the special case where D is a division algebra of degree 8.

Proposition 6.1 ([1, Satz 5.6]). *Let ψ be a quadratic F-form of dimension ≥ 5. The group $H^3(F(\psi)/F)$ is non-trivial iff ψ is a neighbor of an anisotropic 3-Pfister form.*

Proposition 6.2 (see [28, Prop. 4.1 and Rem. 4.1]). *Let D be a central division F-algebra of exponent 2. Suppose that D is decomposable (in the tensor product of two proper subalgebras). Then $H^3(F(D)/F) = [D] \cup H^1(F)$.*

Proposition 6.3. *If D and D' are Brauer equivalent central simple F-algebras, then the function fields $F(D)$ and $F(D')$ are stably equivalent.*[2]

[2] Two field extensions E/F and E'/F are called *stably equivalent*, if some finitely generated purely transcendental extension of E is isomorphic (over F) to some finitely generated purely transcendental extension of E'.

Proof. Since the algebras $D_{F(D')}$ and $D'_{F(D)}$ are split, the field extensions

$$F(D,D')/F(D') \quad \text{and} \quad F(D,D')/F(D)$$

are purely transcendental. Therefore each of the field extensions $F(D)/F$ and $F(D')/F$ is stably equivalent to the extension $F(D,D')/F$. □

Corollary 6.4. *Fix a quadratic F-form ψ and integers $i,j \in \mathbb{Z}$. For any central simple F-algebra D, the groups $H^i(F(D)/F)$, $H^i(F(D)/F, \mathbb{Q}/\mathbb{Z}(j))$, $H^i(F(\psi,D)/F)$, $H^i(F(\psi,D)/F, \mathbb{Q}/\mathbb{Z}(j))$ only depend on the Brauer class of D.* □

Proposition 6.5. *Let D be a central simple F-algebra of exponent 2 and let ψ be a quadratic F-form. The group $H^3(F(\psi,D)/F, \mathbb{Q}/\mathbb{Z}(2))$ is annihilated by 2.*

Proof. Let ψ_0 be a 3-dimensional subform of ψ. Clearly,

$$H^3(F(\psi,D)/F, \mathbb{Q}/\mathbb{Z}(2)) \subset H^3(F(\psi_0,D)/F, \mathbb{Q}/\mathbb{Z}(2)) .$$

Therefore, it suffices to show that the latter cohomology group is annihilated by 2. Replacing ψ_0 by the quaternion algebra $C_0(\psi_0)$, we come to a statement covered by [7, Lemma A.8]. □

Corollary 6.6. *In the conditions of Proposition 6.5, one has*
$$H^3(F(\psi,D)/F, \mathbb{Q}/\mathbb{Z}(2)) = H^3(F(\psi,D)/F) .$$ □

Proposition 6.7. *Let D be a central simple F-algebra of exponent 2 and let ψ be a quadratic F-form of dimension ≥ 3. Suppose that* $\operatorname{ind} D_{F(\psi)} < \operatorname{ind} D$. *Then ψ is not a 3-Pfister neighbor and there is an isomorphism*

$$\frac{H^3(F(\psi,D)/F)}{H^3(F(\psi)/F) + [D] \cup H^1(F)} \simeq \operatorname{Tors}\operatorname{CH}^2(X_\psi \times X_D) .$$

Proof. By [9, Prop. 2.2], there is an isomorphism

$$\frac{H^3(F(\psi,D)/F, \mathbb{Q}/\mathbb{Z}(2))}{H^3(F(\psi)/F, \mathbb{Q}/\mathbb{Z}(2)) + H^3(F(D)/F, \mathbb{Q}/\mathbb{Z}(2))} \simeq$$
$$\simeq \frac{\operatorname{Tors}\operatorname{CH}^2(X_\psi \times X_D)}{pr_\psi^* \operatorname{Tors}\operatorname{CH}^2(X_\psi) + pr_D^* \operatorname{Tors}\operatorname{CH}^2(X_D)} .$$

By Corollary 6.6, we have $H^3(F(\psi,D)/F, \mathbb{Q}/\mathbb{Z}(2)) = H^3(F(\psi,D)/F)$; by [9, Lemma 2.8], we have $H^3(F(\psi)/F, \mathbb{Q}/\mathbb{Z}(2)) = H^3(F(\psi)/F)$; and by [7, Lemma A.8], we have $H^3(F(D)/F, \mathbb{Q}/\mathbb{Z}(2)) = H^3(F(D)/F)$.

Let D' be a division algebra Brauer equivalent to D. By Corollary 6.4, we have $H^3(F(D)/F) = H^3(F(D')/F)$; by [18, Prop. 1.1], we have $\operatorname{Tors}\operatorname{CH}^2(X_D) \simeq \operatorname{Tors}\operatorname{CH}^2(X_{D'})$. Since $D'_{F(\psi)}$ is no more a skew field, there is a homomorphism of F-algebras $C_0(\psi) \to D'$ ([34, Th. 1], see also [26, Th. 2]). Although the algebra $C_0(\psi)$ is not always central simple, it always contains a non-trivial subalgebra central simple over F. Therefore, D' is decomposable, what implies $H^3(F(D')/F) = [D] \cup H^1(F)$ (Proposition 6.2) and $\operatorname{Tors}\operatorname{CH}^2(X_{D'}) = 0$ ([17, Prop. 5.3]). Finally, the existence of a homomorphism $C_0(\psi) \to D'$ implies that ψ is not a 3-Pfister neighbor; therefore $\operatorname{Tors}\operatorname{CH}^2(X_\psi) = 0$ ([14, Th. 6.1]). □

Corollary 6.8. *Let D be a central division F-algebra of degree 8 and exponent 2. Let ψ be a 5-dimensional quadratic F-form. Suppose that $D_{F(\psi)}$ is not a skew field. Then $H^3(F(\psi, D)/F) = [D] \cup H^1(F)$.*

Proof. It is a direct consequence of Proposition 6.7, Corollary 5.6, and Proposition 6.1. □

Theorem 6.9. *Theorem 0.1 is true if D is a division algebra of degree 8.*

Proof. Let ψ_0 be a 5-dimensional subform of ψ. Applying Corollary 6.8, we have $[D] \cup H^1(F) \subset H^3(F(\psi, D)/F) \subset H^3(F(\psi_0, D)/F) = [D] \cup H^1(F)$. Hence $H^3(F(\psi, D)/F) = [D] \cup H^1(F)$.

The assertion on $\operatorname{Tors}\operatorname{CH}^2(X_\psi \times X_D)$ is Corollary 5.6. □

Corollary 6.10. *Let $\phi \in I^2(F)$ be a 8-dimensional quadratic form such that $\operatorname{ind} C(\phi) = 8$. Let D be a degree 8 central simple algebra such that $c(\phi) = [D]$. Let ψ be a quadratic form of dimension ≥ 5 such that $\phi_{F(\psi)}$ is isotropic. Then*

1) *D is a division algebra;*
2) *$D_{F(\psi)}$ is not a division algebra;*
3) *$H^3(F(\psi, D)/F) = [D] \cup H^1(F)$.* □

7. Proof of Corollaries 0.2 and 0.3

We need several lemmas.

Lemma 7.1. *Let $\phi \in I^2(F)$ be a 8-dimensional quadratic form and let D be an algebra such that $c(\phi) = [D]$. Then $\phi_{F(D)} \in GP_3(F(D))$.*

Proof. We have $c(\phi_{F(D)}) = c(\phi)_{F(D)} = [D_{F(D)}] = 0$. Hence $\phi_{F(D)} \in I^3(F(D))$. Since $\dim \phi = 8$, we are done by the Arason-Pfister Hauptsatz. □

Lemma 7.2. *Let $\phi, \phi^* \in I^2(F)$ be 8-dimensional quadratic forms such that $c(\phi) = c(\phi^*) = [D]$, where D is a triquaternion division algebra.*[3] *Suppose that there is a quadratic form ψ of dimension ≥ 5 such that the forms $\phi_{F(\psi,D)}$ and $\phi^*_{F(\psi,D)}$ are isotropic. Then ϕ and ϕ^* are half-neighbors.*

Proof. Lemma 7.1 implies that $\phi_{F(\psi,D)}, \phi^*_{F(\psi,D)} \in GP_3(F(\psi, D))$. By the assumption of the lemma, $\phi_{F(\psi,D)}$ and $\phi^*_{F(\psi,D)}$ are isotropic. Hence $\phi_{F(\psi,D)}$ and $\phi^*_{F(\psi,D)}$ are hyperbolic. Thus $\phi, \phi^* \in W(F(\psi, D)/F)$.

Let $\tau = \phi \perp \phi^*$. Clearly $\tau \in W(F(\psi, D)/F)$. Since $c(\tau) = c(\phi) + c(\phi^*) = [D] + [D] = 0$, we have $\tau \in I^3(F)$. Thus $e^3(\tau) \in H^3(F(\psi, D)/F)$. It follows from Corollary 6.10 that $e^3(\tau) \in [D] \cup H^1(F)$. Hence there exists $s \in F^*$ such that $e^3(\tau) = [D] \cup (s)$. We have $e^3(\tau) = [D] \cup (s) = c(\phi) \cup (s) = e^3(\phi \langle\langle s \rangle\rangle)$. Since $\ker(e^3 : I^3(F) \to H^3(F)) = I^4(F)$, we have $\tau \equiv \phi \langle\langle s \rangle\rangle \pmod{I^4(F)}$. Therefore $\phi + \phi^* = \tau \equiv \phi \langle\langle s \rangle\rangle = \phi - s\phi \pmod{I^4(F)}$. Hence $\phi^* + s\phi \in I^4(F)$. Hence ϕ and ϕ^* are half-neighbors. □

The following statement was pointed out by Laghribi ([23]) as an easy consequence of the index reduction formula [25].

[3] An F-algebra is called *triquaternion*, if it is isomorphic to a tensor product of three quaternion F-algebras.

Lemma 7.3. *Let ψ be a quadratic form of dimension ≥ 5 and D be a division triquaternion algebra. Suppose that $D_{F(\psi)}$ is not a division algebra. Then there exists an 8-dimensional quadratic form $\phi^* \in I^2(F)$ such that $\psi \subset \phi^*$ and $c(\phi^*) = [D]$.* □

Proof of Corollary 0.2. Let D be triquaternion algebra such that $c(\phi) = [D]$. Since $\operatorname{ind} C(\phi) = 8$, it follows that D is a division algebra. Since $\phi_{F(\psi)}$ is isotropic, $D_{F(\psi)}$ is not a division algebra. It follows from Lemma 7.3 that there exists an 8-dimensional quadratic form $\phi^* \in I^2(F)$ such that $\psi \subset \phi^*$ and $c(\phi^*) = [D]$. Obviously, all conditions of Lemma 7.2 hold. Hence ϕ and ϕ^* are half-neighbors. □

Lemma 7.4. *Let D be a division triquaternion algebra over F. Then there exist a field extension E/F and an 8-dimensional quadratic form $\phi^* \in I^2(E)$ with the following properties:*

(i) *D_E is a division algebra,*
(ii) *$c(\phi^*) = [D_E]$,*
(iii) *$\phi^*_{E(D)}$ is anisotropic.*

Proof. Let $\phi \in I^2(F)$ be an arbitrary F-form such that $c(\phi) = [D]$. Let $K = F(X,Y,Z)$ and $\gamma = \phi_K \perp \langle\langle X,Y,Z\rangle\rangle$ be a K-form. Let $K = K_0, K_1, \dots, K_h$; $\gamma_0, \gamma_1, \dots, \gamma_h$ be a generic splitting tower of γ.

Since $\gamma \equiv \phi_K \pmod{I^3(K)}$, we have $c(\gamma) = c(\phi_K) = [D_K]$. Since K/F is purely transcendental, $\operatorname{ind} D_K = \operatorname{ind} D = 8$. Hence $\operatorname{ind} C(\gamma) = 8$. It follows from Corollary 1.5 that there exists s such that $\dim \gamma_s = 8$ and $\operatorname{ind} C(\gamma_s) = 8$. We set $E = E_s$, $\phi^* = \gamma_s$.

We claim that the condition (i)–(iii) of the lemma hold. Since $c(\phi^*) = c(\gamma_E) = c(\phi_E) = [D_E]$, condition (ii) holds. Since $[D_E] = c(\phi^*) = c(\gamma_s)$, we have $\operatorname{ind} D_E = \operatorname{ind} C(\gamma_s) = 8$ and thus condition (i) holds.

Now we only need to verify that (iii) holds. Let M_0/F be an arbitrary field extension such that ϕ_{M_0} is hyperbolic. Let $M = M_0(X,Y,Z)$. We have $\gamma_M = \phi_M \perp \langle\langle X,Y,Z\rangle\rangle_M$. Clearly $\langle\langle X,Y,Z\rangle\rangle$ is anisotropic over M. Since ϕ_M is hyperbolic, we have $(\gamma_M)_{an} = \langle\langle X,Y,Z\rangle\rangle_M$ and hence $\dim(\gamma_M)_{an} = 8$. Therefore $\dim(\gamma_M)_{an} = \dim \gamma_s$. By Lemma 1.3, we see that the field extension $ME/M = MK_s/M$ is purely transcendental. Hence $\dim(\gamma_{ME})_{an} = \dim(\gamma_M)_{an} = 8$. Since $(\phi^*_{ME})_{an} = (\gamma_{ME})_{an}$, we see that ϕ^*_{ME} is anisotropic. Since ϕ_M is hyperbolic, it follows that $[D_M] = c(\phi_M) = 0$. Hence $[D_{ME}] = 0$ and therefore the field extension $ME(D)/ME$ is purely transcendental. Hence $\phi^*_{ME(D)}$ is anisotropic. Therefore $\phi^*_{E(D)}$ is anisotropic. □

Lemma 7.5. *Let $\phi, \phi^* \in I^2(F)$ be 8-dimensional quadratic forms such that $c(\phi) = c(\phi^*) = [D]$, where D is a triquaternion division algebra. Suppose that $\phi^*_{F(D)}$ is anisotropic. Then $\phi_{F(D)}$ is anisotropic.*

Proof. Suppose at the moment that $\phi_{F(D)}$ is isotropic. Then letting $\psi \overset{\text{def}}{=} \phi^*$, we see that all conditions of Lemma 7.2 hold. Hence ϕ and ϕ^* are half-neighbors, i.e., there exists $s \in F^*$ such that $\phi^* + s\phi \in I^4(F)$. Therefore $\phi^*_{F(D)} + s\phi_{F(D)} \in I^4(F(D))$. Since $\phi_{F(D)}$ is isotropic, it is hyperbolic and we see that $\phi^*_{F(D)} \in I^4(F(D))$. By the Arason-Pfister Hauptsatz, we see that $\phi^*_{F(D)}$ is hyperbolic. So we get a contradiction to the assumption of the lemma. □

Proposition 7.6. *Let $\phi \in I^2(F)$ be an 8-dimensional quadratic form such that* $\operatorname{ind} C(\phi) = 8$. *Let A be an algebra such that $c(\phi) = [A]$. Then $\phi_{F(A)}$ is anisotropic.*

Proof. Let D be a triquaternion algebra such that $c(\phi) = [D]$. Since $\operatorname{ind} C(\phi) = 8$, D is a division algebra. Let E/F and ϕ^* be such that in Lemma 7.4. All conditions of Lemma 7.5 hold for E, ϕ_E, ϕ^*, and D_E. Therefore $\phi_{E(D)}$ is anisotropic. Hence $\phi_{F(D)}$ is anisotropic. Since $[A] = c(\phi) = [D]$, the field extension $F(A)/F$ is stably isomorphic to $F(D)/F$ (Proposition 6.3). Therefore $\phi_{F(A)}$ is anisotropic. □

Proof of Corollary 0.3. Suppose at the moment that $\phi_{F(A)} \in I^4(F(A))$. Since $\operatorname{ind} C(\phi) \geq 8$, it follows that $\dim \phi \geq 8$. By Corollary 1.5 there exists a field extension E/F such that $\dim(\phi_E)_{an} = 8$, $\operatorname{ind} C(\phi_E) = 8$. Since $\dim(\phi_E)_{an} = 8$ and $\phi_{E(A)} \in I^4(E(A))$, the Arason-Pfister Hauptsatz shows that $((\phi_E)_{an})_{E(A)}$ is hyperbolic. We get a contradiction to Proposition 7.6. □

8. Proof of Theorem 0.1

By Proposition 6.7, there is a surjection

$$\frac{H^3(F(\psi, D)/F)}{[D] \cup H^1(F)} \twoheadrightarrow \operatorname{Tors} \mathrm{CH}^2(X_\psi \times X_D) \ .$$

Thus, it suffices to prove the second formula of Theorem 0.1.

Proving the second formula, we may assume that $\dim \psi = 5$ (compare to the proof of Theorem 6.9) and D is a division algebra (Corollary 6.4). Under these assumptions, we can write down D as the tensor product $C_0(\psi) \otimes_F B$ (using [25, Th. 1]). In particular, we see that $C_0(\psi)$ is a division algebra, i.e. $\operatorname{ind} C_0(\psi) = \deg C_0(\psi) = 4$.

If $\deg D < 8$, then $D \simeq C_0(\psi)$. In this case, $\psi_{F(D)}$ is a 5-dimensional quadratic form with trivial Clifford algebra; therefore $\psi_{F(D)}$ is isotropic; by this reason, the field extension $F(\psi, D)/F(D)$ is purely transcendental and consequently $H^3(F(\psi, D)/F(D)) = 0$. It follows that

$$H^3(F(\psi, D)/F) = H^3(F(D)/F) = [D] \cup H^1(F) \ ,$$

where the last equality holds by Proposition 6.2.

If $\deg D > 8$, then $\operatorname{ind} B \geq 4$. Applying the index reduction formula [31, Th. 1.3], we get

$$\operatorname{ind} C_0(\psi)_{F(D)} = \min\{\operatorname{ind} C_0(\psi),\ \operatorname{ind} B\} = 4 \ .$$

Therefore $\psi_{F(D)}$ is not a 3-Pfister neighbor and by Proposition 6.1 the group $H^3(F(\psi, D)/F(D))$ is trivial. Thus once again

$$H^3(F(\psi, D)/F) = H^3(F(D)/F) = [D] \cup H^1(F) \ .$$

Finally, if $\deg D = 8$, then we are done by Theorem 6.9 and Proposition 6.7.

References

[1] Arason, J. Kr. *Cohomologische Invarianten quadratischer Formen.* J. Algebra 36 (1975), 448–491.

[2] Esnault, H., Kahn, B., Levine, M., and Viehweg, E. *The Arason invariant and mod 2 algebraic cycles.* J. Amer. Math. Soc., to appear.

[3] Fulton, W. *Intersection Theory.* Springer-Verlag, 1984.

[4] Hartshorne, R. *Algebraic Geometry.* Springer-Verlag, 1977.

[5] Hoffmann, D. W. *Splitting patterns and invariants of quadratic forms.* Math. Nachr., to appear.

[6] Hurrelbrink, J., Rehmann, U. *Splitting patterns of quadratic forms.* Math. Nachr. 176 (1995), 111–127.

[7] Izhboldin, O. T. *On the nonexcellence of the function fields of Severi-Brauer varieties.* Max-Planck-Institut für Mathematik in Bonn, Preprint MPI 96-159 (1996), 1–28.

[8] Izhboldin, O. T., Karpenko, N. A. *Isotropy of virtual Albert forms over function fields of quadrics.* Prépublications de l'Équipe de Mathématique de Besançon 97/07 (1997), 1–11.

[9] Izhboldin, O. T., Karpenko, N. A. *Isotropy of 6-dimensional quadratic forms over function fields of quadrics.* Prépublications de l'Équipe de Mathématique de Besançon 97/12 (1997), 1–25.

[10] Izhboldin, O. T., Karpenko, N. A. *Isotropy of 8-dimensional quadratic forms over function fields of quadrics.* K-Theory Preprint Archives (http://www.math.uiuc.edu/K-theory/), Preprint N°219, 1997.

[11] Izhboldin, O. T., Karpenko, N. A. *Some new examples in the theory of quadratic forms.* K-Theory Preprint Archives (http://www.math.uiuc.edu/K-theory/), Preprint N°234, 1997.

[12] Kahn, B. *Descente galoisienne et K_2 des corps de nombres.* K-Theory 7 (1993), no. 1, 55–100.

[13] Kahn, B. *Motivic cohomology of smooth geometrically cellular varieties.* K-Theory Preprint Archives (http://www.math.uiuc.edu/K-theory/), Preprint N°218, 1997.

[14] Karpenko, N. A. *Algebro-geometric invariants of quadratic forms.* Algebra i Analiz 2 (1991), no. 1, 141–162 (in Russian). Engl. transl.: Leningrad (St. Petersburg) Math. J. 2 (1991), no. 1, 119–138.

[15] Karpenko, N. A. *On topological filtration for Severi-Brauer varieties.* Proc. Symp. Pure Math. 58.2 (1995), 275–277.

[16] Karpenko, N. A. *On topological filtration for Severi-Brauer varieties II.* Transl. Amer. Math. Soc. 174 (1996), no. 2, 45–48.

[17] Karpenko, N. A. *Codimension 2 cycles on Severi-Brauer varieties.* Prépublications de l'Équipe de Mathémathiques de Besançon 96/40 (1996), 1–26. To appear in K-Theory.

[18] Karpenko, N. A. *Cycles de codimension 2 en produits de variétés de Severi-Brauer.* Publications Mathématiques de la Faculté des Sciences de Besançon — Théorie des Nombres, Années 1994/95–1995/96, 1–15.

[19] Karpenko, N. A. *Cohomology of relative cellular spaces and isotropic flag varieties.* Preprint, 1997 (see http://www.uni-muenster.de/math/u/scharlau/publ).

[20] Karpenko, N. A., Merkurjev, A. S. *Chow groups of projective quadrics.* Algebra i Analiz 2 (1990), no. 3, 218–235 (in Russian). Engl. transl.: Leningrad (St. Petersburg) Math. J. 2 (1991), no. 3, 655–671.

[21] Knebusch, M. *Generic splitting of quadratic forms I.* Proc. London Math. Soc. 33 (1976), 65–93.

[22] Knebusch, M. *Generic splitting of quadratic forms II.* Proc. London Math. Soc. 34 (1977), 1–31.

[23] Laghribi, A. *Formes quadratiques en 8 variables dont l'algèbre de Clifford est d'indice 8.* K-Theory, to appear.

[24] Lam, T. Y. *The Algebraic Theory of Quadratic Forms.* Massachusetts: Benjamin 1973 (revised printing: 1980).

[25] Merkurjev, A. S. *Simple algebras and quadratic forms.* Izv. Akad. Nauk SSSR Ser. Mat. 55 (1991), 218–224 (in Russian). English transl.: Math. USSR Izv. 38 (1992), no. 1, 215–221.

[26] Merkurjev, A. S. *K-theory of simple algebras.* Proc. Symp. Pure Math. 58.1 (1995), 65–83.

[27] Panin, I. A. *On the algebraic K-theory of twisted flag varieties.* K-Theory 8 (1994), no. 6, 541–585.

[28] Peyre, E. *Products of Severi-Brauer varieties and Galois cohomology.* Proc. Symp. Pure Math. 58.2 (1995), 369–401.

[29] Quillen, D. *Higher algebraic K-theory I.* Lect. Notes Math. 341 (1973), 85–147.

[30] Scharlau, W. *Quadratic and Hermitian Forms.* Springer, Berlin, Heidelberg, New York, Tokyo (1985).

[31] Schofield, A., van den Bergh, M. *The index of a Brauer class on a Brauer-Severi variety.* Transactions Amer. Math. Soc. 333 (1992), no. 2, 729–739.

[32] Suslin, A. A. *Algebraic K-theory and the norm residue homomorphism.* J. Soviet Math. 30 (1985), 2556–2611.

[33] Swan, R. *K-theory of quadric hypersurfaces.* Ann. Math. 122 (1985), no. 1, 113–154.

[34] Tignol, J.-P. *Réduction de l'indice d'une algèbre simple centrale sur le corps des fonctions d'une quadrique.* Bull. Soc. Math. Belgique 42 (1990), 725–745.

Oleg Izhboldin
Department of
Mathematics and Mechanics
St.-Petersburg
State University
Petrodvorets, 198904
Russia
oleg@izh.usr.pu.ru

Nikita Karpenko
Westfälische
Wilhelms-Universität
Mathematisches Institut
Einsteinstraße 62
D-48149 Münster
Germany
karpenk@math.uni-muenster.de

Remarks on the Darboux Transform of Isothermic Surfaces

Udo Hertrich-Jeromin[1] and Franz Pedit[2]

Received: December 4, 1996
Revised: September 25, 1997

Communicated by Ursula Hamenstädt

Abstract. We study Darboux and Christoffel transforms of isothermic surfaces in Euclidean space. Using quaternionic calculus we derive a Riccati type equation which characterizes all Darboux transforms of a given isothermic surface. Surfaces of constant mean curvature turn out to be special among all isothermic surfaces: their parallel surfaces of constant mean curvature are Christoffel and Darboux transforms at the same time. We prove — as a generalization of Bianchi's theorem on minimal Darboux transforms of minimal surfaces — that constant mean curvature surfaces in Euclidean space allow ∞^3 Darboux transforms into surfaces of constant mean curvature. We indicate the relation between these Darboux transforms and Bäcklund transforms of spherical surfaces.

1991 Mathematics Subject Classification: (Primary) 53A10, (Secondary) 53A50, 53C42.

Keywords: Isothermic surface, Darboux transformation, Christoffel transformation, Riccati equation, Constant mean curvature, Baecklund transformation.

1 Introduction

Transformations play an important role connecting surface theory with the theory of integrable systems. A well known example is the Bäcklund transformation of pseudospherical (and spherical [1]) surfaces in Euclidean 3-space which "adds solitons" to a given surface. In case of isothermic surfaces the Darboux transformation takes the role of the Bäcklund transform for pseudospherical surfaces. Darboux transforms of isothermic surfaces naturally arise in 1-parameter families ("associated families")

[1]Partially supported by the Alexander von Humboldt Stiftung and by NSF grant DMS 93-12087.
[2]Partially supported by NSF grant DMS 93-12087.

allowing to rewrite the underlying (system of) partial differential equation(s) as an (infinite dimensional) integrable system [6], [4]. It is mainly for this reason that Darboux transformations provoke new interest among contemporary geometers — even though the subject was well studied around the turn of the century [5], [7] and [2]. A key tool in the study of Darboux transforms of an isothermic surface in Euclidean space is a careful analysis of the Christoffel transform (or dual isothermic surface) of the surface — which may be considered as a certain limiting case of Darboux transforms. In the present paper, we develop classical results further using quaternionic calculus which makes definitions elegant and calculations more efficient. Characterizations thus obtained turned out to be necessary in the development of the corresponding discrete theory [10].

In the first part of the paper, we develop isothermic surface theory in codimension 2 — which is a more appropriate setting when using quaternionic calculus. When restricting to codimension 1, all notions become classical. Here, we rely on the characterizations of Darboux and Christoffel pairs in $\mathbb{H}P^1$ given in [9]. The consequent use of the quaternionic setup yields a new and unified description for these surface pairs in $\mathbb{R}^4 \cong \mathbb{H}$. Even though the quaternionic calculus (as developed in [9]) provides a setting to study the global geometry of surface pairs in Möbius geometry (cf.[11]) we will restrict to local geometry in this paper, for two reasons: first, there are a number of possible definitions of a "globally isothermic surface" whose consequences have not yet been worked out. For example, definition 1 may well be read as a global definition but it is far too general to provide any global results. Secondly, Christoffel and Darboux transforms of a (compact) surface generally do not exist globally. Moreover, around certain types of umbilics they may not even exist locally. However, up to the problem of closing periods, the results on constant mean curvature surfaces can well be read as global results: here, the Christoffel transform can be determined without integration which ensures its global existence (with branch points at the umbilics of the original surface).

A central result is obtained by carefully analyzing the relation between Darboux and Christoffel pairs: we derive a Riccati type equation describing all Darboux transforms of a given isothermic surface. This equation is crucial for the explicit calculation of Darboux transforms — in the smooth case (all the pictures shown in this paper are obtained from this equation) as well as in the theory of discrete isothermic nets [10]. Moreover, most of our remaining results are different applications of the Riccati equation: first, we extend Bianchi's permutability theorems for Darboux and Christoffel transforms for the codimension 2 setup. We then discuss constant mean curvature surfaces in 3-dimensional Euclidean space as "special" isothermic surfaces: they can be characterized by the fact that their Christoffel transforms arise as Darboux transforms[3]. Together with the Riccati equation, this provides more detailed knowledge about the ∞^3 constant mean curvature Darboux transforms of a constant mean curvature surface — whose existence is a classical result due to Bianchi [1]. Our new proof shows that any such Darboux transform has (pointwise) constant distance to the Christoffel transform. This fact provides a geometric definition for a

[3] While the notion of "Darboux pairs" is naturally a conformal notion (i.e. relates surfaces in Möbius space) the notion of "Christoffel pairs" is a Euclidean one. This might explain the (untypical) fact that constant mean curvature surfaces in *Euclidean* space have a special position, *not* constant mean curvature surfaces in *any* space of constant curvature.

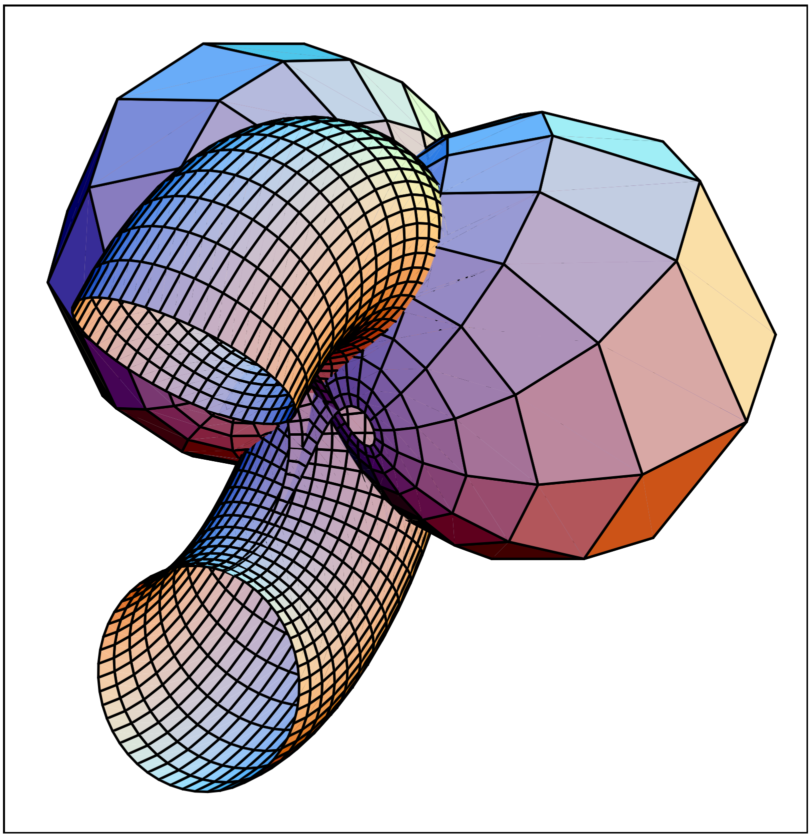


Figure 1: A Darboux transform of a torus of revolution

discrete analog of smooth constant mean curvature surfaces [10]. We conclude this paper relating this 3-dimensional family with the Bianchi-Bäcklund transformation for constant mean curvature surfaces discussed in [12] (cf.[1]).

2 DARBOUX PAIRS IN THE CONFORMAL 4-SPHERE

In 3-dimensional Möbius space (the conformal sphere S^3) an isothermic surface may be characterized by the existence of conformal curvature line coordinates around each (nonumbilic) point[4]. Note that the notion of principal curvature directions is conformally invariant — even though the second fundamental form is not. In higher codimensions the second fundamental form (with respect to any metric in the conformal class) takes values in the normal bundle. In order to diagonalize this vector valued second fundamental form, i.e. simultaneously diagonalize all components of a

[4] As mentioned earlier, there is a variety of possible definitions for isothermic surfaces which are all equivalent away from umbilics — for example, any of the characterizations of isothermic surfaces (cf.[9]) given in this paper could be used as (global) definitions instead of definition 1 (cf.[11]).

basis representation, the surface's normal bundle has to be flat[5]. This is an implicit prerequisite in the following

DEFINITION 1 *A (2-dimensional) surface in (4-dimensional) Möbius space is called isothermic if around each (nonumbilic) point there exist conformal curvature line coordinates, i.e. conformal coordinates which diagonalize the (vector valued) second fundamental form taken with respect to any conformal metric of the ambient space.*

In order to understand the notion of a "Darboux pair of isothermic surfaces" we also have to learn what a "sphere congruence" is and what we will mean by "envelope of a sphere congruence":

DEFINITION 2 *A congruence of 2-spheres in (4-dimensional) Möbius space is a 2-parameter family of 2-spheres.*

A 2-dimensional surface is said to envelope a congruence of 2-spheres if at each point it is tangent[6] to a corresponding 2-sphere.

Note that the requirements on a congruence of 2-spheres in 4-space to be enveloped by two surfaces are much more restrictive than on a hypersphere congruence [9]. Also, a congruence of 2-spheres in S^4 may have only *one* envelope — which generically does not occur in the hypersphere case. In the second half of the paper we will concentrate on the more familiar situation in 3-space.

If, however, we *have* two surfaces which envelope a congruence of 2-spheres the congruence will establish a point to point correspondence between its two envelopes by assigning the point of contact on one surface to the point of contact on the other surface. For a 3-dimensional ambient space it is well known [3] (cf. [7]) that two cases can occur if this correspondence preserves curvature lines[7] and is conformal: the congruence consists of planes in a certain space of constant curvature — in which case the two envelopes are Möbius equivalent — or, both envelopes are isothermic — in this case one surface is called a "Darboux transform" of the other (see [9], compare [3] or [4]). These remarks may motivate the following

DEFINITION 3 *If a congruence of 2-spheres (which is not a plane congruence in a certain space of constant curvature) is enveloped by two isothermic surfaces, the correspondence between its two envelopes being conformal and curvature line preserving, the surfaces are said to form a Darboux pair. Each of the two surfaces is called a Darboux transform of the other.*

Before studying Darboux pairs in Euclidean space we will recall

3 A BASIC CHARACTERIZATION FOR DARBOUX PAIRS

In order to discuss (Darboux) pairs of surfaces in 4- (or 3-) dimensional Möbius geometry we consider the conformal 4-sphere as the quaternionic projective line [9]:

$$S^4 \cong \mathbb{H}P^1 = \{x \cdot \mathbb{H} \mid x \in \mathbb{H}^2\}. \tag{1}$$

[5] Since the principal directions of the (scalar) second fundamental forms with respect to any normal vector are conformally invariant, as in the codimension 1 case, the flatness of the normal bundle is a conformal invariant, too.

[6] As usually done in the 3-dimensional case, we also want to allow the surface to degenerate.

[7] This is what is called a "Ribaucour sphere congruence".

Note that we consider the space $\boldsymbol{H}^2$ of homogeneous coordinates of the quaternionic projective line as a *right* vector space over the quaternions $\boldsymbol{H}$.

Now, let $(f,\hat f): M^2 \to \mathcal{P}$ be an immersion into the (symmetric) space of point pairs[8] in S^4,

$$\mathcal{P} := \{(x,y) \in S^4 \times S^4 \,|\, x \neq y\}. \tag{2}$$

We may write the derivatives of f and $\hat f$ as[9]

$$df = f\varphi + \hat f\omega, \;\; d\hat f = f\hat\omega + \hat f\hat\varphi \tag{3}$$

where $\varphi, \omega, \hat\varphi, \hat\omega : TM \to \boldsymbol{H}$ denote suitable quaternionic valued 1-forms. Then, the integrability conditions $d^2 f = d^2\hat f = 0$ for f and $\hat f$ — the Maurer Cartan equations — read

$$\begin{array}{rcll} 0 & = & d\varphi + \varphi\wedge\varphi + \hat\omega\wedge\omega & \text{(Gauß equation for } f\text{)}, \\ 0 & = & d\omega + \omega\wedge\varphi + \hat\varphi\wedge\omega & \text{(Codazzi equation for } f\text{)}, \\ 0 & = & d\hat\omega + \hat\omega\wedge\hat\varphi + \varphi\wedge\hat\omega & \text{(Codazzi equation for } \hat f\text{)}, \\ 0 & = & d\hat\varphi + \hat\varphi\wedge\hat\varphi + \omega\wedge\hat\omega & \text{(Gauß equation for } \hat f\text{)}. \end{array} \tag{4}$$

Since the quaternions are *not* commutative $\varphi\wedge\varphi \neq 0$ in general. Before continuing, let us list some useful identities for quaternionic 1-forms: let $\alpha, \beta : TM \to \boldsymbol{H}$ be quaternionic valued 1-forms and $g : M \to \boldsymbol{H}$ be a quaternionic valued function; then

$$\begin{array}{rcl} \alpha\wedge g\beta & = & \alpha g\wedge\beta, \\ \overline{\alpha\wedge\beta} & = & -\bar\beta\wedge\bar\alpha, \\ d(g\alpha) & = & dg\wedge\alpha + g\cdot d\alpha, \\ d(\alpha g) & = & -\alpha\wedge dg + d\alpha\cdot g, \end{array} \tag{5}$$

where $(\alpha\wedge\beta)(x,y) := \alpha(x)\beta(y) - \alpha(y)\beta(x)$.

In this framework we are now able to state a basic characterization for Darboux pairs of isothermic surfaces (for more details[10] including a proof see [9]):

PROPOSITION 1 *A pair of surfaces $(f,\hat f): M^2 \to \mathcal{P}$ is a Darboux pair if and only if*

$$\omega\wedge\hat\omega = \hat\omega\wedge\omega = 0 \tag{6}$$

where $\omega, \hat\omega : TM \to \boldsymbol{H}$ are defined by

$$df = f\varphi + \hat f\omega, \;\; d\hat f = f\hat\omega + \hat f\hat\varphi. \tag{7}$$

It is easy to see that this characterization does not depend upon the choice of homogeneous coordinates for the two surfaces: given a change of homogeneous coordinates $(f,\hat f) \mapsto (fa, \hat f\hat a)$, $a, \hat a : M \to \boldsymbol{H}$, we have

$$\begin{array}{rcl} d(fa) & = & (fa)\cdot(a^{-1}\varphi a + a^{-1}da) + (\hat f\hat a)\cdot(\hat a^{-1}\omega a), \\ d(\hat f\hat a) & = & (fa)\cdot(a^{-1}\hat\omega\hat a) + (\hat f\hat a)\cdot(\hat a^{-1}\hat\varphi\hat a + \hat a^{-1}d\hat a). \end{array} \tag{8}$$

---

[8] The homogeneous coordinates of a pair of (different) points in $\boldsymbol{H}P^1$ form a basis of $\boldsymbol{H}^2$. Thus, $\mathcal{P}$ can be identified with the symmetric space $\frac{Gl(2,\boldsymbol{H})}{\boldsymbol{H}_*\times\boldsymbol{H}_*}$. Sometimes it is more convenient to use suitably normalized coordinates: the group $Gl(2,\boldsymbol{H})$ may be replaced by a 15-dimensional subgroup $Sl(2,\boldsymbol{H})$ which is a double cover of the group of orientation preserving Möbius transformations of S^4 [9].

[9] We will use "f" and "$\hat f$" for the point maps into S^4 as well as for their homogeneous coordinates.

[10] In fact, this proposition states the connection between Darboux pairs and "curved flats" [8] in the symmetric space of point pairs.

4 Christoffel pairs of isothermic surfaces in Euclidean space

Another observation is that introducing a real parameter into the Maurer Cartan equations (4) we can obtain the Darboux pair equations (6) together with the original integrability conditions as integrability conditions of a 1-parameter family of Darboux pairs — the "associated family" of Darboux pairs[11]: writing

$$df_r = f_r\varphi + \hat{f}_r(r^2\omega),\ d\hat{f}_r = f_r(r^2\hat{\omega}) + \hat{f}_r\hat{\varphi} \tag{9}$$

with a parameter $r \in \mathbb{R}$ the Gauß equations for f_r and $\hat{f}_r$ become

$$\begin{array}{rcl} 0 & = & d\varphi + \varphi\wedge\varphi + r^4\cdot\hat{\omega}\wedge\omega, \\ 0 & = & d\hat{\varphi} + \hat{\varphi}\wedge\hat{\varphi} + r^4\cdot\omega\wedge\hat{\omega} \end{array} \tag{10}$$

while the Codazzi equations remain unchanged. This shows that if there exist surface pairs — not necessarily Darboux — $(f_r, \hat{f}_r)$ for more than one value of $r > 0$, then, we have a whole 1-parameter family of Darboux pairs.

Assuming we have such a 1-parameter family $(f_r, \hat{f}_r)$ of Darboux pairs a special situation will occur when $r \to 0$. To discuss this, we assume $\varphi = \hat{\varphi} = 0$ without loss of generality: we have $0 = d\varphi + \varphi\wedge\varphi$ and $0 = d\hat{\varphi} + \hat{\varphi}\wedge\hat{\varphi}$ and thus at least locally $\varphi = -da\,a^{-1}$ and $\hat{\varphi} = -d\hat{a}\,\hat{a}^{-1}$ with suitable functions $a, \hat{a} : M \to \mathbb{H}$. Rescaling by those and applying (8) gives $\varphi = \hat{\varphi} = 0$. Thus,

$$df_r = \hat{f}_r(r^2\omega),\ d\hat{f}_r = f_r(r^2\hat{\omega}), \tag{11}$$

and after the rescaling $(f, \hat{f}) \mapsto (f\frac{1}{r}, \hat{f}r)$ (or $(f, \hat{f}) \mapsto (fr, \hat{f}\frac{1}{r})$, respectively) we see that $\hat{f}$ (or f) becomes a fixed point in the conformal 4-sphere — which should be interpreted as a point at infinity. Thus, the other limit surfaces, f_0 and $\hat{f}_0$, naturally lie in (different) Euclidean spaces. Identifying these two Euclidean spaces "correctly" we obtain $df_0 = \bar{\omega}$ and $d\hat{f}_0 = \hat{\omega}$ [9].

These two limit surfaces $\hat{f}_0^c := f_0$ and $f_0^c := \hat{f}_0$ usually do *not* form a Darboux pair — in general they do *not* even envelope a congruence of 2-spheres[12]. But they do form what is called a Christoffel pair:

Definition 4 *Two surfaces $f_0, \hat{f}_0 : M^2 \to \mathbb{R}^4 \cong \mathbb{H}$ in Euclidean 4-space are said to form a Christoffel pair if they induce conformally equivalent metrics on M and have parallel tangent planes with opposite orientations. Each of the surfaces of a Christoffel pair is called a Christoffel transform or dual of the other.*

Note that the two surfaces of a Christoffel pair are automatically isothermic; in fact, isothermic surfaces can be characterized by the (local) existence of a Christoffel transform [9]. The Christoffel transform of an isothermic surface is unique[13] up to a

[11] As we mentioned in a previous footnote (10) Darboux pairs are actually curved flats in the symmetric space of point pairs — and curved flats arise in associated families.

[12] This might seem more natural if we remember that f_0 and $\hat{f}_0$ take values in "different" Euclidean spaces (cf. [4]). — However, one of these surfaces and the point at infinity (which are the remains of the other surface) *do* form a (degenerate) Darboux pair.

[13] Except in one case: Christoffel transforms of the 2-sphere appear in 1-parameter families. We will discuss this case later (see page 324).

homothety and a translation — so that in the sequel we will denote the Christoffel transform of an isothermic surface f by f^c.

Finally, let us state a characterization of Christoffel pairs similar to that for Darboux pairs:

PROPOSITION 2 *Two surfaces $f_0, \hat{f}_0 : M^2 \to \mathbb{R}^4 \cong \mathbb{H}$ form a Christoffel pair if and only if*

$$d\bar{f}_0 \wedge d\hat{f}_0 = d\hat{f}_0 \wedge d\bar{f}_0 = 0. \tag{12}$$

Both surfaces of a Christoffel pair are isothermic.

As for the characterization of Darboux pairs (page 317) a proof may be found in [9]. However, in case of 3-dimensional ambient space we will present an easy proof later (page 323) using some of the calculus we are going to develop.

Now we are prepared to study

5 DARBOUX PAIRS IN $\mathbb{R}^4$

Let $(f, \hat{f}) : M^2 \to \mathcal{P}$ denote a pair of surfaces with

$$df = f\varphi + \hat{f}\omega, \;\; d\hat{f} = f\hat{\omega} + \hat{f}\hat{\varphi}, \tag{13}$$

as before. Assuming that $f, \hat{f} : M \to \mathbb{H} \times \{1\} \cong \mathbb{H}$ take values in Euclidean 4-space we see that $\varphi = -\omega$ and $\hat{\varphi} = -\hat{\omega}$, and hence

$$df = (\hat{f} - f) \cdot \omega, \;\; d\hat{f} = (f - \hat{f}) \cdot \hat{\omega}. \tag{14}$$

This allows us to rewrite condition (6) on f and $\hat{f}$ to form a Darboux pair[14] as

$$0 = df \wedge (f - \hat{f})^{-1} d\hat{f} = d\hat{f} \wedge (\hat{f} - f)^{-1} df. \tag{15}$$

As a first consequence of these equations we derive the equations

$$\begin{array}{rcccl} 0 & = & df \wedge (\hat{f} - f)^{-1} d\hat{f} (\hat{f} - f)^{-1} & = & (\hat{f} - f)^{-1} d\hat{f} (\hat{f} - f)^{-1} \wedge df, \\ 0 & = & d\hat{f} \wedge (f - \hat{f})^{-1} df (f - \hat{f})^{-1} & = & (f - \hat{f})^{-1} df (f - \hat{f})^{-1} \wedge d\hat{f} \end{array} \tag{16}$$

for any Darboux pair $(f, \hat{f})$. Since (15) also implies

$$0 = d[(\hat{f} - f)^{-1} d\hat{f} (\hat{f} - f)^{-1}] = d[(f - \hat{f})^{-1} df (f - \hat{f})^{-1}] \tag{17}$$

we conclude that the Christoffel transforms f^c and $\hat{f}^c$ of f and $\hat{f}$ are given by

$$\begin{array}{rcl} df^c & = & \overline{(\hat{f} - f)^{-1} d\hat{f} (\hat{f} - f)^{-1}}, \\ d\hat{f}^c & = & \overline{(f - \hat{f})^{-1} df (f - \hat{f})^{-1}}. \end{array} \tag{18}$$

Finally, if we fix the translations of f^c and $\hat{f}^c$ such that

$$\overline{(f^c - \hat{f}^c)} = (f - \hat{f})^{-1} \tag{19}$$

— note that $d\overline{(f - \hat{f})}^{-1} = d(f^c - \hat{f}^c)$ — we learn from the above characterization (15) of Darboux pairs that f^c and $\hat{f}^c$ also form a Darboux pair (cf. [2]):

[14] Hopefully, the reader will forgive our context dependent notation: f and $\hat{f}$ denote points in $\mathbb{H}P^1 \cong S^4$, vectors in $\mathbb{H}^2$ or numbers in $\mathbb{H} \cong \mathbb{R}^4$.

Theorem 1 *If $f, \hat{f} : M^2 \to \mathbb{R}^4$ form a Darboux pair, then, their Christoffel transforms $f^c, \hat{f}^c : M^2 \to \mathbb{R}^4$ (if correctly scaled and positioned) form a Darboux pair, too.*

So far we learned how to derive the Christoffel transforms f^c and $\hat{f}^c$ of two surfaces f and $\hat{f}$ forming a Darboux pair. But usually it will be much easier to determine an isothermic surface's Christoffel transform than a Darboux transform. In the next section we will see that deriving Darboux transforms $\hat{f}$ and $\hat{f}^c$ of two surfaces f and f^c forming a Christoffel pair[15] comes down to solving

6 A Riccati type equation

Solving (18) for $d\hat{f}$ we obtain $d\hat{f} = (\hat{f} - f)d\bar{f}^c(\hat{f} - f)$. This yields the following Riccati type partial differential equation[16] for $g := (\hat{f} - f)$:

$$dg = g\, d\bar{f}^c\, g - df. \tag{20}$$

Using our characterization (12) of Christoffel pairs it is easily seen that this equation is "completely" (Frobenius) integrable. Note that — in agreement with our previous results — the common transform $g^c = \bar{g}^{-1}$ for Riccati equations yields

$$dg^c = g^c\, d\bar{f}\, g^c - df^c, \tag{21}$$

showing that $\hat{f}^c = f^c + g^c$ will provide a Darboux transform of f^c whenever $f + g$ is a Darboux transform of f coming from a solution g of (20).

Since every Darboux transform $\hat{f}$ of an isothermic surface f provides a Christoffel transform f^c of f via (18) *every* Darboux transform comes from a solution of (20) — if we do not fix the scaling of the Christoffel transform f^c. On the other hand *every* solution g of (20) defines a Darboux transform $\hat{f} = f + g$ of f since $df \wedge g^{-1}d(f+g) = d(f+g) \wedge g^{-1}df = 0$. This seems to be worth formulating as a

Theorem 2 *If $f, f^c : M^2 \to \mathbb{R}^4$ form a Christoffel pair of isothermic surfaces every solution of the integrable Riccati type partial differential equation*

$$dg = g\, d\bar{f}^c\, g - df \tag{22}$$

provides a Darboux transform $\hat{f} = f + g$ of f. On the other hand, every Darboux transform $\hat{f}$ of f is obtained this way — if we do not fix the scaling of f^c.

At this point, we should discuss the effect of a rescaling of the Christoffel transform f^c in the equation (20). For this purpose we examine the equations

$$dg = g\,(\pm r^4 d\bar{f}^c)\, g - df \tag{23}$$

[15] Note that the notation $\hat{f}^c$ for a Darboux transform of f^c makes sense because of our previous theorem: we have $\widehat{f^c} = \hat{f}^c$.

[16] The pictures in this paper were produced using Mathematica to numerically integrate this Riccati type equation.

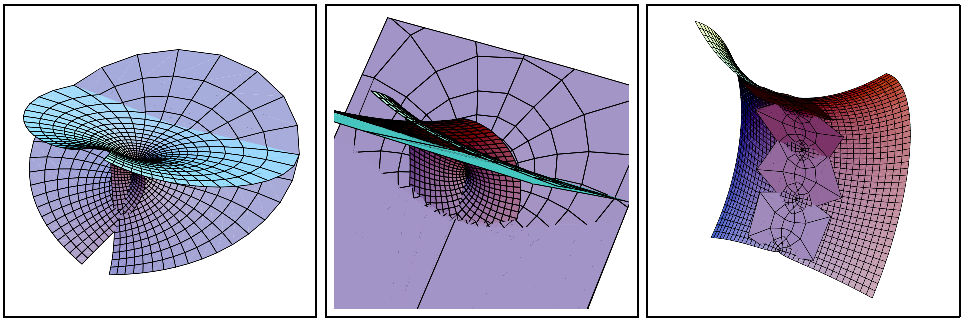


Figure 2: Darboux transforms of the Catenoid when $H^c \to \infty$

where $r \neq 0$ is a real parameter. For the derivatives of f and a Darboux transform $\hat{f} = f + g$ of f this yields

$$\begin{array}{rcl} df &=& f \cdot [-g^{-1}df] \quad + \quad \hat{f} \cdot [g^{-1}df], \\ d\hat{f} &=& f \cdot [\mp r^4 d\bar{f}^c\, g] \quad + \quad \hat{f} \cdot [\pm r^4 d\bar{f}^c\, g]. \end{array} \tag{24}$$

Interpreting $f, \hat{f} : M^2 \to \mathbb{H} \cong \mathbb{H} \times \{1\}$ as homogeneous coordinates of the point pair map $(f, \hat{f}) : M^2 \to \mathcal{P}$ we may choose new homogeneous coordinates by performing a rescaling $(f, \hat{f}) \mapsto (fr, \hat{f}(rg)^{-1})$ to obtain[17]

$$\begin{array}{rcl} d[fr] &=& [fr] \cdot [-g^{-1}df] \quad + \quad [\hat{f}(rg)^{-1}] \cdot [r^2 df], \\ d[\hat{f}(rg)^{-1}] &=& [fr] \cdot [\mp r^2 d\bar{f}^c] \quad + \quad [\hat{f}(rg)^{-1}] \cdot [df\, g^{-1}]. \end{array} \tag{25}$$

Even though this system resembles very much our original system (9) which describes the associated family of Darboux pairs, there is an essential difference: in (9) the forms φ, ω, $\hat{\varphi}$ and $\hat{\omega}$ are independent of the parameter r whereas the forms $g^{-1}df$ and $df\, g^{-1}$ in the system we just derived *do* depend on r. In fact, in the associated family $(f_r, \hat{f}_r)$ of Darboux pairs obtained from (9) both surfaces, f_r as well as $\hat{f}_r$, change with the parameter r whereas the parameter contained in the Riccati equation just effects the Darboux transform $\hat{f} = \hat{f}_r$ while the original surface f remains unchanged. However, the original system (9) appears in the linearization of our Riccati equation[18] which indicates a close relation of these two parameters.

As a first application of this parameter which occurs from rescalings of the Christoffel transform f^c in our Riccati equation we may prove an extension of Bianchi's permutability theorem [2] for Darboux transforms:

THEOREM 3 *Let $\hat{f}_{1,2} : M^2 \to \mathbb{H}$ be two Darboux transforms of an isothermic surface $f : M^2 \to \mathbb{H}$,*

$$d\hat{f}_{1,2} = r_{1,2}(\hat{f}_{1,2} - f)\, d\bar{f}^c(\hat{f}_{1,2} - f), \tag{26}$$

where we fixed any scaling for the Christoffel transform f^c of f. Then, there exists an isothermic surface $\hat{\hat{f}} : M^2 \to \mathbb{H}$ which is an r_1-Darboux transform of $\hat{f}_2$ and an

[17] Note that this rescaling provides an $Sl(2, \mathbb{H})$ framing of the point pair map $(f, \hat{f})$ [9].
[18] Here, we would like to thank Fran Burstall for helpful discussions.

r_2-Darboux transform of $\hat f_1$ at the same time[19]*:*

$$d\hat f = r_{2,1}(\hat f - \hat f_{1,2})\, d\bar{\hat f}^{c}_{1,2}(\hat f - \hat f_{1,2}). \tag{27}$$

Moreover, the points of $\hat f$ lie on the circles determined by the corresponding points of f, $\hat f_1$ and $\hat f_2$, the four surfaces having a constant (real) cross ratio[20]

$$\frac{r_2}{r_1} \equiv (f - \hat f_1)(\hat f_1 - \hat f)^{-1}(\hat f - \hat f_2)(\hat f_2 - f)^{-1}. \tag{28}$$

To prove this theorem we simply *define* the surface $\hat f : M^2 \to \mathbb{H}$ by solving the cross ratio equation[21] (28) for $\hat f$:

$$\hat f := [r_2 \hat f_1 (\hat f_1 - f)^{-1} - r_1 \hat f_2 (\hat f_2 - f)^{-1}] \cdot [r_2(\hat f_1 - f)^{-1} - r_1(\hat f_2 - f)^{-1}]^{-1}. \tag{29}$$

Using this ansatz, it is a straightforward calculation to verify the Riccati equations (27) which proves the theorem.

As indicated earlier, from now on we will concentrate on surfaces in 3-dimensional Euclidean space $\mathbb{R}^3 \cong \mathrm{Im}\mathbb{H}$:

7 Christoffel pairs in $\mathbb{R}^3$

In this situation, much of our previously developed calculus will simplify considerably. For example, we will be able to give an easy proof of our characterization of Christoffel pairs and to write down the Christoffel transform of an isothermic surface quite explicitly. First we note that our characterizations (15) and (12) of Darboux and Christoffel pairs of isothermic surfaces reduce to just *one* equation: if $f, \hat f : M^2 \to \mathrm{Im}\mathbb{H}$ both take values in the imaginary quaternions,

$$\begin{array}{rcl} d\hat f \wedge d\bar f &=& -\overline{d\bar f \wedge d\hat f}, \\ d\hat f \wedge (\hat f - f)^{-1} df &=& -\overline{df \wedge (f - \hat f)^{-1} d\hat f}. \end{array} \tag{30}$$

In order to continue we will collect some identities present in the codimension 1 case. We may orient an immersion $f : M^2 \to \mathbb{R}^3 \cong \mathrm{Im}\mathbb{H}$ by choosing a unit normal field $n : M^2 \to S^2$. This defines the complex structure J on M via

$$df \circ J = n\, df \tag{31}$$

— note that since f and n take values in the imaginary quaternions

$$n\, df = -\langle n, df\rangle + n \times df = n \times df = -df\, n. \tag{32}$$

The Hodge operator is then given as the dual of this complex structure:

$$*\eta = -\eta \circ J \tag{33}$$

[19] Note, that this claim makes no sense before we fix a scaling for the Christoffel transforms $\hat f^c_{1,2}$ of $\hat f_{1,2}$. But, according to our "permutability theorem" for Christoffel and Darboux transforms (theorem 1) there is a *canonical* scaling for $\hat f^c_{1,2}$ after we fixed the scaling of f^c.

[20] For a comprehensive discussion of the (complex) cross ratio in $\mathbb{R}^4 \cong \mathbb{H}$ see [10]. The idea for the proof given in this paper actually originated from the discrete version of this theorem.

[21] Note that the denominator does not vanish as long as $\hat f_1 \neq \hat f_2$. For $r_1 = r_2$ we get $\hat f = f$.

for any 1-form η on M.

With this notation we are now able to give a useful reformulation[22] of the equation arising in our characterizations of Darboux pairs and Christoffel pairs: if $\eta : TM \to \mathbb{H}$ is any quaternionic valued 1-form we have

$$(df \wedge \eta)(x, Jx) = df(x) \cdot (-*\eta(x) + n\,\eta(x)) \tag{34}$$

for any $x \in TM$. Consequently, $df \wedge \eta = 0$ if and only if

$$*\eta = n\,\eta. \tag{35}$$

This criterium shows that the space of imaginary solutions $\eta : TM \to \mathrm{Im}\mathbb{H}$ of the equation $0 = df \wedge \eta$ is pointwise 2-dimensional[23] — if η is an (injective) solution, then, every other solution $\tilde{\eta}$ is of the form

$$\tilde{\eta} = (a + b\,n) \cdot \eta \tag{36}$$

with suitable functions $a, b : M \to \mathbb{R}$. But *one* (imaginary) solution to the equation $0 = df \wedge \eta$ is easily found: it is well known that

$$d * df = -dn \wedge df = H\,df \wedge df \tag{37}$$

where H is the mean curvature of f. Thus

$$df \wedge (dn + H\,df) = 0 \tag{38}$$

which gives an injective solution $\eta = dn + H\,df$ away from umbilics of f.

At this point, we are ready to give the announced proof of our characterization of Christoffel pairs (12) in the 3-dimensional case:

THEOREM 4 *Two surfaces $f, f^c : M^2 \to \mathbb{R}^3 \cong \mathrm{Im}\mathbb{H}$ form a Christoffel pair if and only if*

$$df \wedge df^c = 0. \tag{39}$$

Generically, the Christoffel transform f^c of f is uniquely determined by f up to homotheties and translations of $\mathbb{R}^3$.

The fact that both surfaces of a Christoffel pair in 3-space are isothermic is classical (see for example [5]) — and thus we omit this calculation.

Now, in order to prove this theorem we note that from the above we know that $f^c : M^2 \to \mathrm{Im}\mathbb{H}$ satisfies (39) if and only if

$$*df^c = n\,df^c. \tag{40}$$

[22] At this point we would like to thank Ulrich Pinkall for many helpful discussions — this criterium is actually due to him.

[23] The space of solutions with values in the full quaternions is 4-dimensional as is easily seen: (36) becomes

$$\tilde{\eta} = (a + b\,n) \cdot \eta + (*\alpha + n\alpha)$$

with an arbitrary real 1-form $\alpha : TM \to \mathbb{R}$.

But this equation means that in corresponding points f and f^c have parallel tangent planes and that the almost complex structure induced by f^c with respect to $n^c := -n$ is just J — the same as that induced by f with respect to n. Thus,

$$df \wedge df^c = 0 \tag{41}$$

if and only if $f, f^c : M^2 \to \mathbb{R}^3$ have parallel tangent planes with opposite orientations and they induce conformally equivalent metrics, i.e. they form a Christoffel pair.

Now assume we have not just one but two Christoffel transforms f^c and $\tilde{f}^c$ of an isothermic surface $f : M^2 \to \mathbb{R}^3$. Then we know from (36) that

$$d\tilde{f}^c = (a + b\,n) \cdot df^c. \tag{42}$$

The integrability condition for $\tilde{f}^c$ reads

$$0 = da \wedge df^c + db \wedge *df^c + b\,H^c df^c \wedge df^c \tag{43}$$

showing that $a = const$ and $b = 0$ since $df^c \wedge df^c$ takes values in normal direction while all other components are tangential — provided that f^c is *not* a minimal surface[24]. This concludes the proof.

With (38) it also follows that

$$dn + H\,df = (a + b\,n)df^c \tag{44}$$

for suitable functions $a, b : M \to \mathbb{R}$. Similarly, we obtain

$$-dn + H^c\,df^c = (a^c + b^c n)df \tag{45}$$

by interchanging the roles of f and f^c. Adding these two equations yields $a = H^c$, $a^c = H$ and $b = b^c = 0$ since the forms df, $n\,df$, df^c and $n\,df^c$ are linearly independent (over the reals). As a consequence, we have a quite explicit formula relating the two surfaces of a Christoffel pair:

$$H^c df^c = dn + H\,df. \tag{46}$$

This equation shows that whenever one of the surfaces of a Christoffel pair is a minimal surface the other is totally umbilic (namely, a scaling of its Gauß map) and vice versa. This brings us back to our previous problem of the uniqueness of Christoffel transforms: assume we have a Christoffel pair (f, n) consisting of a minimal surface f and its Gauß map n. Then all the pairs

$$(\,a \int (\cos(t) + \sin(t)\,n) \cdot df\,,\ n\,) \tag{47}$$

with real constants a and t will also form Christoffel pairs. Up to homotheties (given by a) this will run us through the associated family of minimal surfaces (given by t) reflecting the fact that associated minimal surfaces have the same Gauß map[25].

Another fact that can be derived from (46) is that the (correctly scaled and positioned) Christoffel transform of a surface of constant mean curvature $H \neq 0$ is its

[24] The case of minimal Christoffel transforms will be discussed below.

[25] However, choosing "curvature lines" for the Gauß map will fix the minimal surface [9].

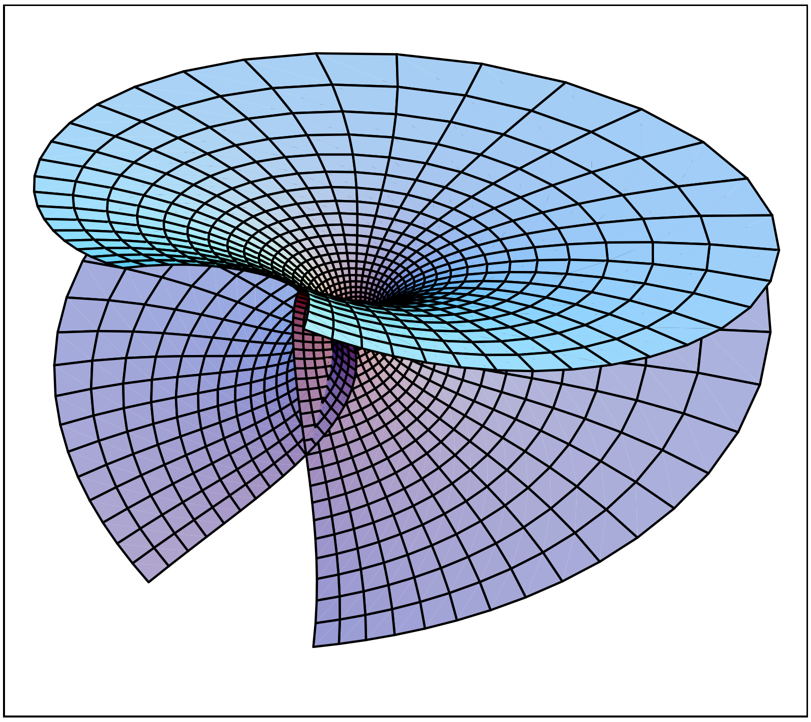


Figure 3: A Darboux transform of the Catenoid

parallel surface $f+\frac{1}{H}n$ of the same constant mean curvature $H^c = H$. Note that this parallel surface induces a conformally equivalent metric on the underlying manifold M^2 and consequently it is also a Darboux transform of the original surface[26] — the enveloped sphere congruence consisting of spheres with constant radius $\frac{1}{2H}$. Later, we will see that constant mean curvature surfaces in Euclidean space can be *characterized* by the fact that their Christoffel transforms are Darboux transforms too. Thus, in the remaining part of this paper we will study constant mean curvature ($H \neq 0$ or $H = 0$) Darboux transforms of

8 Surfaces of constant mean curvature

Using the reformulation (35) of our characterizing equation (15) of Darboux pairs we conclude that for any Darboux transform $\hat{f} = f + g$ of $f : M^2 \to \mathbb{R}^3$

$$*g\,d\hat{f} = n\,g\,d\hat{f} \tag{48}$$

where we used the fact that $g^{-1} = -\frac{1}{|g|^2}g$ for $g \in \mathrm{Im}\mathbb{H}$. Consequently, the normal field $\hat{n}$ of $\hat{f}$ is given by[27]

$$\hat{n} = \frac{gng}{|g|^2} = \frac{1}{|g|^2}(|g|^2 n - 2\langle n, g\rangle g) \tag{49}$$

since we must have $*d\hat{f} = -\hat{n}d\hat{f}$.

Thus, if the normal field of a Darboux transform $\hat{f}$ of an isothermic surface $f : M^2 \to \mathrm{Im}\mathbb{H}$ equals that of its Christoffel transform,

$$\hat{n} = n^c = -n, \tag{50}$$

then $g = a\,n$ for a suitable constant $a \in \mathbb{R}$ (remark that a has to be constant in order to obtain parallel tangent planes of $\hat{f}$ and f). With (46) we conclude

$$Hdf + dn = H^c df^c = H^c(df + dg) = H^c df + H^c a\,dn \tag{51}$$

which implies that either one of the surfaces is minimal and the other is totally umbilic, or, $H = H^c = \frac{1}{a}$ which means that f and $\hat{f} = f^c$ form a pair of parallel constant mean curvature surfaces.

Together with our previous remark (page 326) this leaves us with the following characterization of constant mean curvature surfaces:

---

[26] Note that in order to obtain $g = \frac{n}{H}$ as a solution of our Riccati type equation (20) the Christoffel transform df^c of f has to be scaled such that $H^c = \frac{1}{H}$ — then, the Riccati equation is equivalent to (46). This means that the parallel constant mean curvature surface appears at a well defined location in the associated family.

[27] Note that with this formula we easily see that $\hat{f}$ is the second envelope of a sphere congruence enveloped by f:

$$2\langle g, n\rangle f + |g|^2 n = 2\langle g, n\rangle \hat{f} + |g|^2 \hat{n}.$$

The second fundamental form of $\hat{f}$ is quite complicated, but at least, when introducing frames it can be seen that it has the same principal directions as the second fundamental form of f^c. Since $\hat{f}$ also induces the conformally equivalent metric $|d\hat{f}|^2 = |g|^4|df^c|^2$ we get half of a proof for our characterization (15) of Darboux pairs in the case of 3-dimensional ambient space.

THEOREM 5 *The (correctly scaled and positioned) Christoffel transform f^c of an isothermic surface $f : M^2 \to \mathbb{R}^3$ is also a Darboux transform $\hat{f}$ of f if and only if f is a surface of constant mean curvature $H \neq 0$. In this case $\hat{f} = f^c$ is the parallel surface of constant mean curvature.*

In order to study constant mean curvature Darboux transforms of constant mean curvature surfaces in general we have to calculate the mean curvature of a Darboux transform $\hat{f}$ of an isothermic surface. We will eventually derive the existence of a 3-parameter family of constant mean curvature Darboux transforms of a constant mean curvature surface, all of them having (pointwise) constant distance from the parallel constant mean curvature surface of the original surface. There are several ways to do so: we could calculate the second fundamental form of $\hat{f}$ — which is not convenient because this second fundamental form looks quite difficult — or, we could use (37) to directly calculate $\hat{H}$ with the help of our Riccati type equation (20). This second way is quite straightforward but not very interesting. So, we will present another way which grew out of discussions with Ulrich Pinkall[28]: observing that if $d\hat{f} = -\bar{g}df^c g$, the integrability condition for $\hat{f}$ becomes

$$0 = \bar{g}(\overline{dg\,g^{-1}} \wedge df^c - df^c \wedge dg\,g^{-1})g, \tag{52}$$

i.e. the reality of the form $df^c \wedge dg\,g^{-1}$. Since the volume form $\frac{1}{2}df^c \wedge * \, df^c$ induced by f^c is a basis of the real 2-forms on M this may be reformulated as

$$0 = df^c \wedge (dg\,g^{-1} - \frac{1}{2}U * df^c) \tag{53}$$

with a suitable function $U : M \to \mathbb{R}$. With (35) we obtain the equivalent equation

$$n^c dg - *dg = U\,df^c g \tag{54}$$

— the "Dirac equation" with reference immersion f^c.

Using this equation we may calculate the mean curvature $\hat{H}$ of $\hat{f}$ in terms of the function U via

$$d * d\hat{f} = \frac{1}{|g|^2}(U - H^c)\,d\hat{f} \wedge d\hat{f} \tag{55}$$

since

$$*\alpha \wedge *\beta = \alpha \wedge \beta \tag{56}$$

for any two 1-forms $\alpha, \beta : TM \to \mathbb{H}$ on a Riemann surface and hence

$$*df^c \wedge dg = \frac{1}{2}(*df^c \wedge dg - *df^c \wedge **\,dg) = \frac{1}{2}df^c \wedge (n^c dg - *dg). \tag{57}$$

Substituting our Riccati equation (20) into the Dirac equation yields $U = 2\langle n, g\rangle$ and consequently

$$\hat{H} = \frac{1}{|g|^2}(2\langle n, g\rangle - H^c). \tag{58}$$

[28] The Dirac equation (54) which we will discover on our way can be considered as a replacement for the Cauchy Riemann equations in a generalized "Weierstraß representation" for surfaces in $\mathbb{R}^3$. Given an immersion $f : M^2 \to \mathbb{R}^3$ this generalized "Weierstraß representation" will provide us with *any* immersion $\hat{f}$ which induces the same complex structure on M.

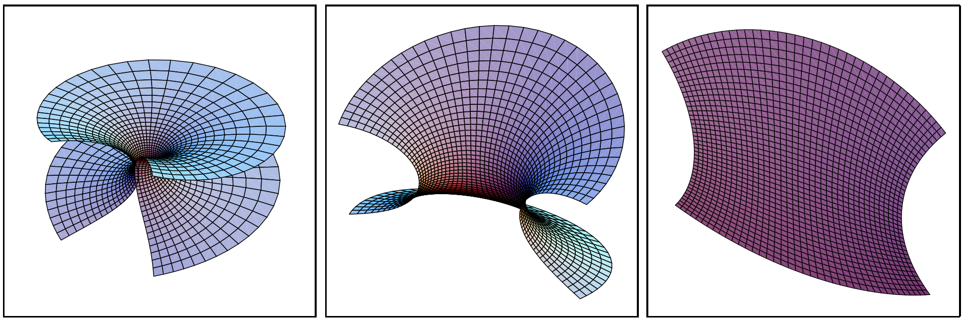


Figure 4: Darboux transforms of the Catenoid when $H^c \to 0$

Now we assume the mean curvature H of our original surface f to be constant — and consequently H^c is constant too — and rewrite this equation as

$$0 = h_{\hat{H}}(g) := \hat{H}\,|g|^2 - 2\langle n, g\rangle + H^c. \tag{59}$$

Taking the derivative of this function h_C where C denotes any constant and assuming H^c to be constant yields

$$dh_C(g) = -2\langle df^c, g\rangle \cdot h_C(g) - 2\langle df, g\rangle \cdot (C - H) \tag{60}$$

where we got rid of dn by using (46). This shows that whenever we choose an initial value $g(p_0) = g_0$ for a function $g : M^2 \to \mathrm{Im}\mathbb{H}$ such that $h_H(g_0) = 0$ the trivial solution $h_H \equiv 0$ will be the unique solution to the above (linear and homogeneous: $C = H$) differential equation. Thus our Riccati type equation (20) will produce a Darboux transform $\hat{f} = f + g$ of constant mean curvature $\hat{H} = H$ out of a surface of constant mean curvature ($H \neq 0$ or $H = 0$).

To conclude let us study the geometry of the condition $h_H(g) = 0$: for a minimal surface this simply says that the points $\hat{f}(p)$ of $\hat{f} = f + g$ always lie in distance $\frac{1}{2}H^c$ off the tangent planes $f(p) + d_pf(T_pM)$ of f. Since we also have the freedom of rescaling the Christoffel transform f^c of f we end up with a 3-parameter family of minimal Darboux transforms of a minimal surface (cf. [2]). A minimal Darboux transform of the Catenoid is shown in figure 3. Sending $H^c \to \pm\infty$ — note that in case of surfaces of constant mean curvature the associated family of Darboux pairs may be parameterized by H^c — the Darboux transforms look more and more like the original surface (Fig. 2) while sending $H^c \to 0$ the Darboux transforms approach a planar surface patch — the best compromise between the Catenoid's Christoffel transform and a minimal surface (Fig. 4).

In case of a surface of constant mean curvature $H \neq 0$ we may reformulate the condition $h_H(g) = 0$ as

$$|H\,g - n|^2 = 1 - H^c H \tag{61}$$

showing that the points $\hat{f}(p)$ lie on spheres centered on the parallel surface $f + \frac{1}{H}n$ and with constant radius $\frac{1}{H}\sqrt{1 - H^c H}$. Since the radius has to be real to provide real Darboux transforms we see that we have to have $H^c H \leq 1$ which restricts the range of the parameter H^c to a ray $H^c \leq \frac{1}{H}$ containing 0 (without loss of generality we assume

$H \geq 0$). As $H^c \to -\infty$ and $H^c \to 0$ we obtain the original surface and its Christoffel transform, respectively. But now, we obtain the Christoffel transform a second time — as a Darboux transform when $H^c = \frac{1}{H}$, i.e. when the spheres $h_H(g) = 0$ collapse to points. Figures 5 and 6 show constant mean curvature Darboux transforms of the cylinder.

To summarize the results we found in this section we formulate a theorem generalizing Bianchi's theorem on minimal Darboux transforms of minimal surfaces [2]:

THEOREM 6 *Any surface of constant mean curvature ($H \neq 0$ or $H = 0$) in Euclidean 3-space allows a 3-parameter family of Darboux transforms into surfaces of the same constant mean curvature.*

In case of a minimal surface all its minimal Darboux transforms have (pointwise) constant normal distance from the original surface while,

in case of a surface of constant mean curvature $H \neq 0$, all the constant mean curvature Darboux transforms have (pointwise) constant distance from the parallel constant mean curvature surface of the original surface.

Having a second look at the Darboux transform of the cylinder shown in figure 5 we recognize a strong similarity to Ivan Sterling's "doublebubbleton" [12]. This suggests a relation between our constant mean curvature Darboux transform and

9 THE BIANCHI-BÄCKLUND TRANSFORM OF CONSTANT MEAN CURVATURE SURFACES

We may supply any surface $f : M^2 \to \mathbb{R}^3$ of constant mean curvature $H = \frac{1}{2}$ with conformal coordinates $(x, y) : M^2 \to \mathbb{R}^2$ such that

$$\begin{array}{rcl} I &=& e^{2u}(dx^2 + dy^2), \\ II &=& e^u(\sinh(u)dx^2 + \cosh(u)dy^2) \end{array} \tag{62}$$

— reflecting the fact that every surface of constant mean curvature is isothermic. Then, a new surface of constant mean curvature — a "Bianchi-Bäcklund transform" of the original surface — can be obtained as $\hat{f} = f + g$ where

$$g = \frac{2}{\sinh(\beta)\cosh(\beta + \varphi)}\left(\cosh(\beta)e^{-u}[\cos\psi f_x - \sin\psi f_y] - \sinh\varphi n\right), \tag{63}$$

β denoting a real parameter and $\varphi + i\psi = \theta$ being given by the linear system

$$\begin{array}{rcl} \theta_x + iu_y &=& \sinh\beta\sinh\theta\cosh u + \cosh\beta\cosh\theta\sinh u \\ i\theta_y + u_x &=& -\sinh\beta\cosh\theta\sinh u - \cosh\beta\sinh\theta\cosh u. \end{array} \tag{64}$$

In fact, this transformation is obtained by applying two successive Bäcklund transforms to the surface of constant Gauß curvature [1] which is parallel to the original surface of constant mean curvature and then, taking the (correct) parallel surface of constant mean curvature [12]. In this construction, the second Bäcklund transform has to be matched to the first one such that the resulting surface of constant Gauß curvature is a real surface again.

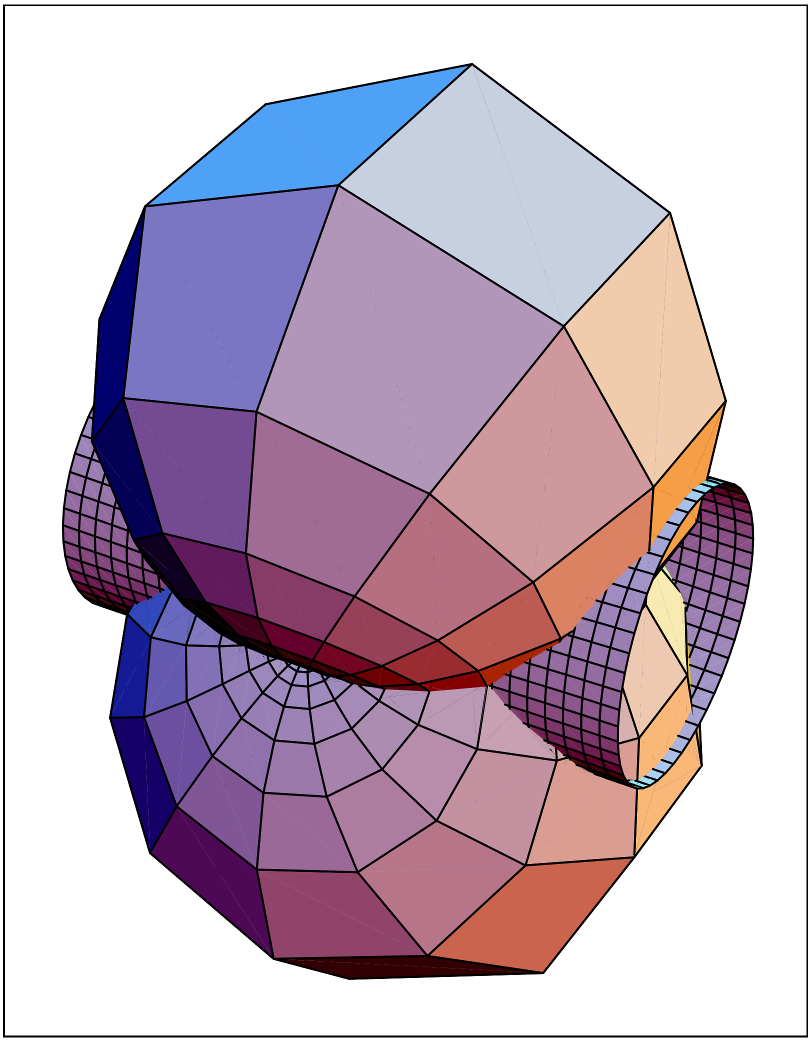


Figure 5: A Darboux transform of the cylinder

Fixing the scaling of the Christoffel transform f^c of f such that $H^c = H = \frac{1}{2}$, i.e. $f^c = f + 2n$, it is an unpleasant but straightforward calculation to see that our Riccati type equation

$$dg = g(\frac{\sinh^2(\beta)}{4} df^c)g - df \tag{65}$$

is equivalent to the above linear system (64) defining the function θ. Thus we have:

THEOREM 7 *Any Bianchi-Bäcklund transform of a surface of constant mean curvature is a Darboux transform.*

Analyzing the effect of the three parameters (β and initial values for φ and ψ) contained in the Bianchi-Bäcklund transform on the function $g : M \to \mathbb{R}^3$ at an initial point we find that any solution of our Riccati equation (20) with a *positive* multiple of the parallel constant mean curvature surface $f + 2n$ as Christoffel transform f^c can be obtained via a Bianchi-Bäcklund transform[29]. Those constant mean curvature Darboux transforms of a constant mean curvature surface where the Christoffel transform is taken a *negative* multiple of the parallel constant mean curvature surface (see Fig. 6) seem *not* to occur as Bianchi-Bäcklund transforms.

[29] Hereby, we also have to allow singularities $\varphi \to \infty$ to obtain vertical values of g too.

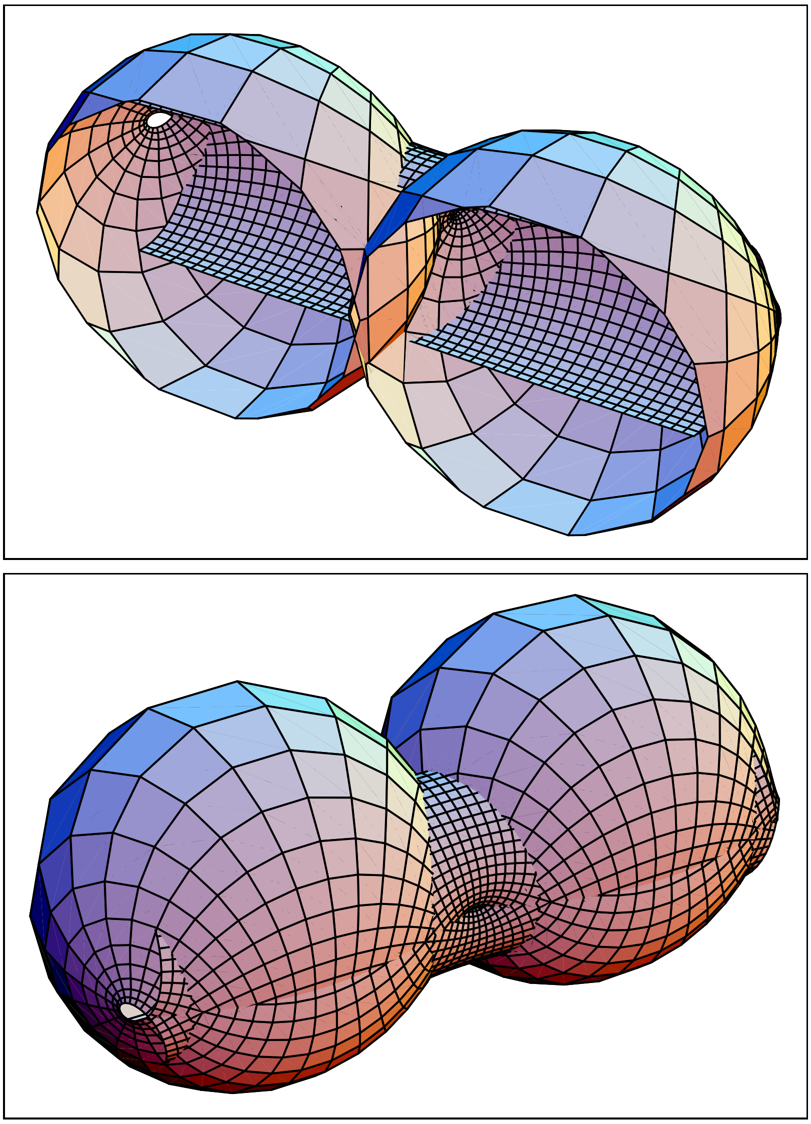


Figure 6: Another Darboux transform of the cylinder

REFERENCES

[1] L. Bianchi: *Vorlesungen über Differentialgeometrie;* Teubner, Leipzig 1899 (Anhang zu Kapitel XVII: Zur Transformationstheorie der Flächen mit constantem positiven Krümmungsmass, pp.641-648)

[2] L. Bianchi: *Ricerche sulle superficie isoterme e sulla deformazione delle quadriche;* Annali Mat. 11 (1905) 93-157

[3] W. Blaschke: *Vorlesungen über Differentialgeometrie III;* Springer, Berlin 1929

[4] F. Burstall, U. Hertrich-Jeromin, F. Pedit and U. Pinkall: *Curved flats and Isothermic surfaces;* Math. Z. 225 199-209

[5] E. Christoffel: *Ueber einige allgemeine Eigenschaften der Minimumsflächen;* Crelle's J. 67 (1867) 218-228

[6] J. Cieśliński: *The Darboux-Bianchi transformation for isothermic surfaces;* Diff. Geom. Appl. 7 (1997) 1-28

[7] G. Darboux: *Sur les surfaces isothermiques;* Comptes Rendus 128 (1899) 1299-1305, 1538

[8] D. Ferus, F. Pedit: *Curved flats in symmetric spaces;* Manuscripta Math. 91 (1996) 445-454

[9] U. Hertrich-Jeromin: *Supplement on Curved flats in the space of point pairs and Isothermic surfaces: A quaternionic Calculus;* Doc. Math. J. DMV 2 (1997) 335-351

[10] U. Hertrich-Jeromin, T. Hoffmann, U. Pinkall: *A discrete version of the Darboux transform for isothermic surfaces;* to appear in A. Bobenko, R. Seiler, *Discrete integrable Geometry and Physics*, Oxford Univ. Press, Oxford 1997

[11] G. Kamberov: *Holomorphic quadratic differentials and conformal immersions;* Preprint (1997)

[12] I. Sterling, H.C. Wente: *Existence and Classification of Constant Mean Curvature Multibubbletons of Finite and Infinite Type;* Indiana Univ. Math. J. 42 (1993) 1239-1266

Udo Hertrich-Jeromin
Dept. Math. & Stat., GANG
University of Massachusetts
Amherst, MA 01003 (USA)
jeromin@math.umass.edu

Franz Pedit
Dept. Math. & Stat., GANG
University of Massachusetts
Amherst, MA 01003 (USA)
pedit@math.umass.edu

Supplement on Curved Flats in the Space of Point Pairs and Isothermic Surfaces: A Quaternionic Calculus

Udo Hertrich-Jeromin*

Received: September, 25, 1997
Revised: December 4, 1997

Communicated by Ursula Hamenstädt

Abstract. A quaternionic calculus for surface pairs in the conformal 4-sphere is elaborated. It introduces a rich algebraic structure and allows the use of global frames while, at the same time, incorporates the classical "geometric" model of Möbius geometry providing geometric clarity. This way, it provides the foundation for the development of new techniques in Möbius differential geometry.

A field where the quaternionic calculus already proved particularly useful is the geometry of transformations of isothermic surfaces: in the second half of the paper, the relation of Darboux and Christoffel pairs of isothermic surfaces and curved flats in the symmetric space of point pairs is discussed and some applications are sketched. In particular, a new viewpoint on relations between surfaces of constant mean curvature in certain spaces of constant curvature, and on Bryant's Weierstrass type representation for surfaces of constant mean curvature 1 in hyperbolic 3-space is presented.

1991 Mathematics Subject Classification: (Primary) 53A10, (Secondary) 53A50, 53C42.

Keywords: Isothermic surface, Darboux transformation, Christoffel transformation, Goursat transformation, Curved flat, Constant mean curvature, Weierstrass representation.

* Partially supported by the Alexander von Humboldt Stiftung and NSF Grant DMS93-12087.

1. Introduction

It is well known that the orientation preserving Möbius transformations of the "conformal 2-sphere" $S^2 \cong \mathbb{C} \cup \{\infty\}$ can be described as fractional linear transformations $z \mapsto \frac{a_{11}z+a_{12}}{a_{21}z+a_{22}}$ where $a = (a_{ij}) \in Sl(2,\mathbb{C})$. The reason for this fact is that the conformal 2-sphere $S^2 \cong \mathbb{C}P^1$ can be identified with the complex projective line. Introducing homogeneous coordinates $p = v_p\mathbb{C}$, $v_p \in \mathbb{C}^2$, on $\mathbb{C}P^1$ the special linear group $Sl(2,\mathbb{C})$ acts on $\mathbb{C}P^1$ by projective transformations — which are, for 1-dimensional projective spaces, identical with Möbius transformations — via $v_p\mathbb{C} \mapsto Av_p\mathbb{C} = v_q\mathbb{C}$. Thus, in affine coordinates one has

$$\begin{pmatrix} z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} z \\ 1 \end{pmatrix} \simeq \begin{pmatrix} \frac{a_{11}z+a_{12}}{a_{21}z+a_{22}} \\ 1 \end{pmatrix}.$$

This (algebraic) model of Möbius geometry in dimension 2 complements the ("geometric") model commonly used in differential geometry: here, the conformal 2-sphere (or, more general, the conformal n-sphere) is considered as a quadric in the real projective 3-space $\mathbb{R}P^3$ and the group of Möbius transformations is isomorphic to the group of projective transformations of $\mathbb{R}P^3$ that map the "absolute quadric" S^2 onto itself (cf.[3]). Equipping the space of homogeneous coordinates of $\mathbb{R}P^3$ with a Lorentz scalar product that has the points of S^2 as isotropic (null) lines, the Möbius group can be identified with the pseudo orthogonal group of this Minkowski space $\mathbb{R}^4_1$.

Several attempts have been made to generalize the described algebraic model to higher dimensions — in particular to dimensions 3 and 4, by using quaternions (cf.[14],[15]): analogous to the above model, the conformal 4-sphere is identified with the quaternionic projective line, $S^4 \cong \mathbb{H}P^1$, with $Sl(2,\mathbb{H})$ acting on it by Möbius transformations. In order to use such an "algebraic model" in Möbius *differential* geometry, it is not enough to describe the underlying space and the Möbius group acting on it, though. One also needs a convenient description for (hyper-) spheres since the geometry of surfaces in Möbius geometry is often closely related to the geometry of an enveloped sphere congruence (cf.[3]). For example, Willmore surfaces in S^3 can be related to minimal surfaces in the space of 2-spheres in S^3, and the geometry of isothermic surfaces is related to that of "sphere surfaces" with flat normal bundle, "Ribaucour sphere congruences".

One way is to identify a hypersphere $s \subset \mathbb{H}P^1$ with the inversion at this sphere. The problem with this approach is, that only the orientation preserving Möbius transformations are naturally described in the algebraic model — but, inversions are orientation reversing Möbius transformations. Adjoining the (quaternionic) conjugation as a basic orientation reversing Möbius transformation and working with the larger group of *all* Möbius transformations, works relatively fine for 2-dimensional Möbius geometry, but turns into a nightmare[1)] in dimension 4 since the quaternions form a non commutative field.

Another way is to identify a sphere $s \subset S^4 \cong \mathbb{H}P^1$ with that quaternionic hermitian form on the space $\mathbb{H}^2$ of homogeneous coordinates that has this sphere s as a

[1)] Identifying 2-spheres in $S^3 \subset S^4 \cong \mathbb{H}P^1$ with inversions in S^4 provides a solution in the codimension 1 case, though: as the composition of two inversions at hyperspheres, the inversion at a 2-sphere in S^4 is orientation preserving.

null cone. After discussing some basics in quaternionic linear algebra we will follow this approach — to obtain not only a description for the space of spheres but also to establish the relation with the classical "geometric" model of Möbius geometry: the space of quaternionic hermitian forms will canonically turn into a real six dimensional Minkowski space, the classical model space.

This (second) way, we combine the advantages of both models for Möbius differential geometry: on one side, we introduce a rich algebraic structure which provides a significant simplification of calculations and, at the same time, we also obtain a calculus that will be more suitable to discuss the global geometry of surfaces in Möbius geometry, as well as the geometry of discrete nets. On the other side, we keep a close connection to the classical model of Möbius geometry which will make it easier to understand the results geometrically. In particular, our calculus will provide an ideal setting for the study of surface *pairs*, maps into the (symmetric) space of point pairs in $\mathbb{H}P^1$ — in Möbius differential geometry, surfaces often occur naturally in pairs, as envelopes of certain distinguished sphere congruences: for example, Willmore surfaces come in dual pairs as envelopes of their common central sphere congruences, and isothermic surfaces permit pairings via Darboux (and Christoffel) transforms.

The latter will be examined in the remaining part of the paper, on one side to see the calculus at work, on the other side to demonstrate some new results: here, our quaternionic calculus provides very elegant characterizations for Darboux and Christoffel pairs of isothermic surfaces that led to the discovery of the Riccati type equation (cf.[11]) for the Darboux transformation of isothermic surfaces — an equation that apparently cannot be derived in the classical calculus (cf.[2]). This is one reason, why the presented calculus was necessary to develop the definition of the discrete version of the Darboux transformation for discrete isothermic nets and the (geometric) definition of discrete cmc nets (cf.[10]). The mentioned characterizations rely on the relation between Darboux pairs of isothermic surfaces and curved flats in the space of point pairs — since this space will turn out to be symmetric the notion of curved flats makes sense. Although this relation was already established in [6] for the codimension 1 case, it might be of interest to see that it also holds in the higher codimension case[2)] of Darboux pairs in $\mathbb{H}P^1$ (cf.[13]). Even though our calculus also provides a framework to discuss global aspects of isothermic surfaces (cf.[12]) we will only focus on their local geometry: there is a variety of possible definitions of "globally isothermic surfaces" whose degree of generality and whose consequences are yet to be worked out. However, computer experiments seem to indicate that Darboux (and Christoffel) transforms of isothermic surfaces only exist locally, in general. And, worse, near certain types of umbilics even their local existence is not clear — resp. depends on the chosen definition of a "globally isothermic surface" ...

In the last section, we study minimal and constant mean curvature surfaces in 3-dimensional spaces of constant curvature. These are "special" isothermic surfaces, and a suitable Christoffel transform in $\mathbb{R}^3$ can be determined algebraically (in the general case, an integration has to be carried out). Examining the effect of the spectral parameter that comes with a curved flat, we obtain a new interpretation for the relations between surfaces of constant curvature in certain space forms. In fact, these

2) Most recently, these results were generalized to arbitrary codimension using an extension of the presented calculus [5].

relations can be interpreted in terms of Bianchi's "T-transformation" for isothermic surfaces [2]. For example, the well known relation between minimal surfaces in the (metric) 3-sphere and surfaces of constant mean curvature in Euclidean space, as well as the relation between minimal surfaces in Euclidean 3-space and surfaces of constant mean curvature 1 in hyperbolic 3-space are discussed. In case of the constant mean curvature 1 surfaces in hyperbolic 3-space, a new form of Bryant's Weierstrass type representation [4] is given. In this context, the classical Enneper-Weierstrass representation for minimal surfaces in Euclidean 3-space is described as a Goursat type transform of the (multiply covered) plane — similar to the way certain surfaces of constant Gauss curvature are described as a Bäcklund transforms of a line. Finally, the classical Goursat transformation for minimal surfaces is generalized for isothermic surfaces in Euclidean space.

2. THE STUDY DETERMINANT

Throughout this paper we will use various well known models [1] for the non commutative field of quaternions:

$$\begin{array}{rcl} I\!H & \cong & \{a+v \,|\, a\in I\!R\cong \mathrm{Re}I\!H,\, v\in I\!R^3\cong \mathrm{Im}I\!H\} \\ & \cong & \{a_0+a_1i+a_2j+a_3k \,|\, a_0,a_1,a_2,a_3\in I\!R\} \\ & \cong & \{x+y\,j \,|\, x,y\in \mathbb{C}\} \\ & \cong & \{A\in M(2\times 2,\mathbb{C}) \,|\, \mathrm{tr}A\in I\!R,\, A+A^*\in I\!RI\}. \end{array}$$

Herein, we can identify i,j,k with the standard basis vectors of $I\!R^3\cong \mathrm{Im}I\!H$: if $v,w\in \mathrm{Im}I\!H$ are two "vectors" their product $v\,w=-v\cdot w+v\times w$ which is equivalent to the familiar identities $i^2=j^2=k^2=-1$, $ij=k=-ji$, $jk=i=-kj$ and $ki=j=-ik$. Obviously, the first model will turn out particularly useful when focusing on the geometry of 3-space while the decomposition $I\!H\cong \mathbb{C}+\mathbb{C}\,j$ will prove useful in the context of surfaces, 2-dimensional submanifolds, since their tangent planes (and normal planes) carry a natural complex structure. We will switch between these models as it appears convenient.

As the quaternions can be thought of as a Euclidean 4-space, $I\!R^4\cong I\!H$, the (conformal) 4-sphere $S^4\cong I\!R^4\cup\{\infty\}$ can be identified with the quaternionic projective line: $S^4\cong I\!HP^1=\{\text{lines through } 0 \text{ in } I\!H^2\}$. Thus, a point $p\in S^4$ of the conformal 4-sphere is described by its homogeneous coordinates $v_p\in I\!H^2$; and its stereographic projection onto Euclidean 4-space $I\!R^4\cong\{v\in I\!H^2\,|\,v_2=1\}$ is obtained by normalizing the second component of v_p.

Since the quaternions form a non commutative field, we have to agree whether the scalar multiplication in a quaternionic vector space is from the right or left: in this paper, $I\!H^2$ will be considered a *right vector space* over the quaternions. This way, quaternionic linear transformations can be described by the multiplication (of column vectors) with (quaternionic) matrices from the *left*: $A(v\lambda)=(Av)\lambda$. For a quaternionic 2-by-2 matrix $A\in M(2\times 2,I\!H)$ we introduce the Study determinant[3)] [1] (cf. Study's "Nablafunktion" [14])

$$\begin{array}{rcl} \mathcal{D}(A) & := & \det(A^*A) \\ & = & |a_{11}|^2|a_{22}|^2+|a_{12}|^2|a_{21}|^2-(\bar a_{11}a_{12}\bar a_{22}a_{21}+\bar a_{21}a_{22}\bar a_{12}a_{11}). \end{array}$$

[3)] Note, that the notion of determinant makes sense for self adjoint matrices $A\in M(2\times 2,I\!H)$.

This is exactly the determinant of the complex 4-by-4 matrix corresponding to A when using the complex matrix model for the quaternions. Thus, $\mathcal{D}$ clearly satisfies the usual multiplication law, $\mathcal{D}(AB) = \mathcal{D}(A)\mathcal{D}(B)$, and vanishes exactly when A is singular. The multiplication law implies that $\mathcal{D}$ is actually an invariant of the linear transformation described by a matrix: $\mathcal{D}(U^{-1}AU) = \mathcal{D}(A)$ for any basis transformation $U : \mathbb{H}^2 \to \mathbb{H}^2$. Also note that $0 \leq \mathcal{D}(A) \in \mathbb{R}$.

Definition. *The general and special linear groups of $\mathbb{H}^2$ will be denoted by*

$$\begin{array}{rcl} Gl(2,\mathbb{H}) & := & \{A \in M(2\times 2,\mathbb{H}) \,|\, \mathcal{D}(A) \neq 0\} \\ Sl(2,\mathbb{H}) & := & \{A \in M(2\times 2,\mathbb{H}) \,|\, \mathcal{D}(A) = 1\}. \end{array}$$

With the help of Study's determinant, the inverse of a quaternionic 2-by-2 matrix $A \in Gl(2,\mathbb{H})$ can be expressed directly as

$$A^{-1} = \frac{1}{\mathcal{D}(A)} \begin{pmatrix} |a_{22}|^2\bar{a}_{11} - \bar{a}_{21}a_{22}\bar{a}_{12} & |a_{12}|^2\bar{a}_{21} - \bar{a}_{11}a_{12}\bar{a}_{22} \\ |a_{21}|^2\bar{a}_{12} - \bar{a}_{22}a_{21}\bar{a}_{11} & |a_{11}|^2\bar{a}_{22} - \bar{a}_{12}a_{11}\bar{a}_{21} \end{pmatrix}.$$

Note also, that $Sl(2,\mathbb{H})$ is a 15-dimensional Lie group — it will turn out to be a double cover of the identity component of the Möbius group of S^4.

Considering $\mathcal{D} : \mathbb{H}^2 \times \mathbb{H}^2 \to \mathbb{R}$ as a function of two (column) vectors we see that $\mathcal{D}(v, v + w) = \mathcal{D}(v, w)$ and $\mathcal{D}(v, w\lambda) = |\lambda|^2\mathcal{D}(v, w)$ — similar formulas holding for the first entry since $\mathcal{D}$ is symmetric: $\mathcal{D}(v, w) = \mathcal{D}(w, v)$. Reformulating our previous statement, we also obtain that $\mathcal{D}(v, w) = 0$ if and only if v and w are linearly dependent[4]. Particularly, if v and w are points in an affine quaternionic line, say the Euclidean 4-space $\{v \in \mathbb{H}^2 \,|\, v_2 = 1\}$, then $\mathcal{D}(v, w) = |v_1 - w_1|^2$ measures the distance between v and w with respect to a Euclidean metric. This fact can be used to express the cross ratio of four points in Euclidean 4-space (cf.[10]) in terms of the Study determinant[5]:

$$|DV(h_1, h_2, h_3, h_4)|^2 = \frac{\mathcal{D}\begin{pmatrix} h_1 & h_2 \\ 1 & 1 \end{pmatrix}\mathcal{D}\begin{pmatrix} h_3 & h_4 \\ 1 & 1 \end{pmatrix}}{\mathcal{D}\begin{pmatrix} h_2 & h_3 \\ 1 & 1 \end{pmatrix}\mathcal{D}\begin{pmatrix} h_4 & h_1 \\ 1 & 1 \end{pmatrix}}.$$

The expression on the right hand is obviously invariant under individual rescalings of the vectors which shows that the cross ratio is, in fact, an invariant of four points in the quaternionic projective line $\mathbb{H}P^1$.

3. Quaternionic hermitian forms

will be a key tool in our calculus for Möbius geometry: any quaternionic hermitian form $s : \mathbb{H}^2 \times \mathbb{H}^2 \to \mathbb{H}$,

$$\begin{array}{rcl} s(v, w_1\lambda + w_2\mu) & = & s(v, w_1)\lambda + s(v, w_2)\mu \\ s(v_1\lambda + v_2\mu, w) & = & \bar{\lambda}s(v_1, w) + \bar{\mu}s(v_2, w) \\ s(w, v) & = & \overline{s(v, w)}, \end{array}$$

[4] All these properties are also easily checked directly, without using the complex matrix representation of the quaternions.

[5] For a more complete discussion of the *complex* cross ratio of four points in space consult [10].

is determined by its values on a basis (e_1, e_2) of $I\!H^2$, $s_{ij} = s(e_i, e_j)$. Since s is hermitian, $s_{11}, s_{22} \in I\!R$ and $s_{21} = \bar{s}_{12} \in I\!H$, the quaternionic hermitian forms on $I\!H^2$ form a 6-dimensional (real) vector space. Clearly, $Gl(2, I\!H)$ operates on this vector space via $(A, s) \mapsto As := [(v, w) \mapsto s(Av, Aw)]$, or, in the matrix representation of s, via $(A, s) \mapsto A^* sA$. A straightforward calculation shows that $\det(As) = \mathcal{D}(A)\det(s)$. This enables us to introduce a Lorentz scalar product

$$\langle s, s \rangle := -\det(s) = |s_{12}|^2 - s_{11}s_{22}$$

on the space $I\!R^6_1$ of quaternionic hermitian forms, which is well defined up to a scale[6)] (or, the choice of a basis in $I\!H^2$). Fixing a scaling of this Lorentz product, the special linear transformations act as isometries on $I\!R^6_1$ — $Sl(2, I\!H)$ is a double cover of the identity component[7)] of $SO_1(6)$, which itself is isomorphic to the group of orientation preserving Möbius transformations of S^4. Thus, restricting our attention to Euclidean 4-space $\{e_1 h + e_2 \,|\, h \in I\!H\}$, the orientation preserving Möbius transformations appear as fractional linear transformations (cf.[14],[15])

$$\begin{pmatrix} h \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} h \\ 1 \end{pmatrix} \simeq \begin{pmatrix} (a_{11}h + a_{12})(a_{21}h + a_{22})^{-1} \\ 1 \end{pmatrix}.$$

If $s \neq 0$ lies in the light cone of $I\!R^6_1$, $\langle s, s \rangle = 0$, then the corresponding quadratic form $v \mapsto s(v, v)$ annihilates exactly *one* direction $vI\!H \subset I\!H^2$: $0 = s(v, v)$ vanishes iff $0 = |s_{11}v_1 + s_{12}v_2|^2$ or $0 = |s_{21}v_1 + s_{22}v_2|^2$ since at least one, s_{11} or s_{22} does not vanish. Hence, we can identify a point $p = vI\!H \in I\!HP^1$ of the quaternionic projective line — the 4-sphere — with the null line of quaternionic hermitian forms in the Minkowski $I\!R^6_1$ that annihilate this point. In homogeneous coordinates, this identification can be given by[8)]

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \leftrightarrow \begin{pmatrix} |v_2|^2 & -v_1\bar{v}_2 \\ -v_2\bar{v}_1 & |v_1|^2 \end{pmatrix} = s_v. \tag{1}$$

Note, that with this identification, $\langle s_v, s \rangle = -s(v, v)$ for any quaternionic hermitian form $s \in I\!R^6_1$. If $s = s_w$ is an isotropic form too, then $\langle s_v, s_w \rangle = -\mathcal{D}(v, w)$.

If, on the other hand, $\langle s, s \rangle = 1$ we obtain — depending on whether $s_{11} = 0$ or $s_{11} \neq 0$ in the chosen basis (e_1, e_2) of $I\!H^2$ —

$$s = \begin{pmatrix} 0 & -n \\ -\bar{n} & 2d \end{pmatrix} \qquad \text{or} \qquad s = \frac{1}{r}\begin{pmatrix} 1 & -m \\ -\bar{m} & |m|^2 - r^2 \end{pmatrix}$$

with suitable n resp. $m \in I\!H$ and d resp. $r \in I\!R$: the null cone of s is a plane with unit normal n and distance d from the origin or a sphere with center m and radius r in Euclidean 4-space $\{e_1 h + e_2 \,|\, h \in I\!H\}$. Consequently, we identify the Lorentz sphere $S^5_1 \subset I\!R^6_1$ with the space of spheres and planes in Euclidean 4-space, or with the space of spheres in S^4 — as the readers familiar with the classical model (cf.[3]) of Möbius geometry might already have suspected. The incidence of a point $p \in S^4 \cong I\!HP^1$ and a sphere $s \subset S^4$, i.e. $s \in S^5_1$, is equivalent to $s(p, p) = 0$ in our quaternionic model.

A key concept in

6) At this point, we notice that the geometrically significant space is the projective 5-space $I\!RP^5$ with absolute quadric $Q = \{I\!Rx \,|\, \langle x, x \rangle = 0\}$, not its space of homogeneous coordinates, $I\!R^6_1$.

7) Using a basis of quaternionic hermitian forms, it is an unpleasant but straightforward calculation to establish a Lie algebra isomorphism $\mathfrak{sl}(2, I\!H) \leftrightarrow \mathfrak{o}_1(6)$.

8) Note the analogy with the Veronese embedding.

4. MÖBIUS DIFFERENTIAL GEOMETRY

is that of (hyper-) sphere congruences and envelopes of sphere congruences:

DEFINITION. *An immersion $f : M \to S^4$ is called an envelope of a hypersphere congruence $s : M \to S^5_1$ if, at each point $p \in M$, f touches the corresponding sphere $s(p)$: $f(p) \in s(p)$ and $d_pf(T_pM) \subset T_{f(p)}s(p)$.*

According to our previous discussion, the first condition — the incidence of $f(p)$ and the corresponding sphere $s(p)$ — is equivalent to $s(f,f) = 0$ in our quaternionic model. Calculating, for a moment, in a Euclidean setting — i.e. $s = \frac{1}{r}\begin{pmatrix} 1 & -m \\ -\bar{m} & |m|^2 - r^2 \end{pmatrix}$ — we find $s(f,df) + s(df,f) = \frac{2}{r}(f - m) \cdot df$. Thus[9],

LEMMA. *An immersion $f : M \to I\!\!H\!P^1$ envelopes a sphere congruence $s : M \to S^5_1$ if and only if $s(f,f) = 0$ and $s(f,df) + s(df,f) = 0$.*

Before going on, we introduce the symmetric space of point pairs: given two (distinct) points of the quaternionic projective line $I\!\!H\!P^1$, we may identify these points with a quaternionic linear transformation P which maps a (fixed) basis (e_1, e_2) of $I\!\!H^2$ to their homogeneous coordinates — or, in coordinates, with a matrix having for columns the homogeneous coordinates of the two points. This linear transformation P is obviously not uniquely determined by the two points in $I\!\!H\!P^1$: any gauge transform $P \cdot H$ of P with H in the isotropy subgroup $K := \{H \in Gl(2, I\!\!H) \mid He_1 = e_1\lambda, He_2 = e_2\mu\}$ determines the same point pair. Thus, the space $\mathcal{P}$ of point pairs in the conformal 4-sphere $I\!\!H\!P^1$ is a homogeneous space, $\mathcal{P} = Gl(2, I\!\!H)/K$. Moreover, the decomposition $\mathfrak{gl}(2, I\!\!H) = \mathfrak{k} \oplus \mathfrak{p}$ with

$$\begin{array}{rcl} \mathfrak{k} & = & \{X \in \mathfrak{gl}(2, I\!\!H) \mid Xe_1 = e_1\lambda, Xe_2 = e_2\mu\} \\ \mathfrak{p} & = & \{X \in \mathfrak{gl}(2, I\!\!H) \mid Xe_1 = e_2\lambda, Xe_2 = e_1\mu\} \end{array} \tag{2}$$

is a Cartan decomposition since $[\mathfrak{k}, \mathfrak{k}] \subset \mathfrak{k}$ $[\mathfrak{k}, \mathfrak{p}] \subset \mathfrak{p}$ and $[\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{k}$ so that $\mathcal{P}$ is, in fact, a symmetric space.

Now, if $F = (f, \hat{f}) : M \to Gl(2, I\!\!H)$ is a framing (lift) of a point pair map $M \to \mathcal{P}$, a simple calculation using (1) shows that

$$Ff = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad F\hat{f} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

if the relative scaling of f and $\hat{f}$ is chosen such that F takes values in the special linear group $Sl(2, I\!\!H)$. Since $Sl(2, I\!\!H)$ acts by isometries on the space $I\!\!R^6_1$ of quaternionic hermitian forms, for any sphere congruence $s : M \to S^5_1$ containing the points of f and $\hat{f}$, we have

$$Fs = \begin{pmatrix} 0 & s_0 \\ \bar{s}_0 & 0 \end{pmatrix}$$

[9] Note, that with the identification (1) of points in $I\!\!H\!P^1$ with isotropic quaternionic hermitian forms, $s(f,df) + s(df,f) = -\langle s, df \rangle$ which gives the link with the classical model of Möbius geometry.

with a suitable function $s_0 : M \to S^3 \subset I\!H$ taking values in the unit quaternions. Passing to another set of homogeneous coordinates by means of a gauge transformation $(f,\hat{f}) \mapsto (f\lambda, \hat{f}\hat{\lambda})$ results in $s_0 \mapsto \bar{\lambda} s_0 \hat{\lambda}$. Thus, depending on a given sphere congruence s, we may fix the homogeneous coordinates of f and $\hat{f}$ such that $s_0 \equiv 1$ — leaving us with a scaling freedom $(f,\hat{f}) \mapsto (f\lambda, \hat{f}\bar{\lambda}^{-1})$ with $\lambda \in I\!H$. A second sphere congruence $\tilde{s}$ (orthogonal to the first one) can be used to further fix the scalings via $\tilde{s}_0 \equiv i$ up to $\lambda \in \mathbb{C}$. Giving a complete set of four accompanying orthogonal sphere congruences and fixing a third one, $\hat{s}$, to satisfy $\hat{s}_0 \equiv j$ leaves us with the familiar real scaling freedom, $\lambda \in I\!R$ (cf.[3]). These choices of accompanying spheres, and accordingly these choices of homogeneous coordinates for a point pair map $(f,\hat{f})$ are the only aspect of the presented calculus that will generally not work globally.

Writing down the derivatives $df = f\varphi + \hat{f}\psi$ and $d\hat{f} = f\hat{\psi} + \hat{f}\hat{\varphi}$ of f and $\hat{f}$, we obtain the connection form

$$\Phi := F^{-1}dF = \begin{pmatrix} \varphi & \hat{\psi} \\ \psi & \hat{\varphi} \end{pmatrix} : TM \to \mathfrak{gl}(2, I\!H)$$

of a framing $F : M \to Gl(2, I\!H)$. A gauge transformation $(f,\hat{f}) \mapsto (f\lambda, \hat{f}\hat{\lambda})$ of the frame will result in a change

$$\begin{pmatrix} \varphi & \hat{\psi} \\ \psi & \hat{\varphi} \end{pmatrix} \mapsto \begin{pmatrix} \lambda^{-1}\varphi\lambda & \lambda^{-1}\hat{\psi}\hat{\lambda} \\ \hat{\lambda}^{-1}\psi\lambda & \hat{\lambda}^{-1}\hat{\varphi}\hat{\lambda} \end{pmatrix} + \begin{pmatrix} \lambda^{-1}d\lambda & 0 \\ 0 & \hat{\lambda}^{-1}d\hat{\lambda} \end{pmatrix} \tag{3}$$

of the connection form Φ. The integrability conditions $0 = d^2 f = d^2\hat{f}$ yield the Maurer-Cartan equation $0 = d\Phi + \Phi \wedge \Phi$ for the connection form: the Gauss-Ricci equations for f resp. $\hat{f}$,

$$\begin{array}{rcl} 0 & = & d\varphi + \varphi\wedge\varphi + \hat{\psi}\wedge\psi \\ 0 & = & d\hat{\varphi} + \hat{\varphi}\wedge\hat{\varphi} + \psi\wedge\hat{\psi}, \end{array} \tag{4}$$

and the Codazzi equations,

$$\begin{array}{rcl} 0 & = & d\psi + \psi\wedge\varphi + \hat{\varphi}\wedge\psi \\ 0 & = & d\hat{\psi} + \hat{\psi}\wedge\hat{\varphi} + \varphi\wedge\hat{\psi}. \end{array} \tag{5}$$

Note, that since the quaternions are not commutative, generally $\varphi\wedge\varphi \neq 0$. Moreover, $d(\lambda\varphi) = d\lambda\wedge\varphi + \lambda d\varphi$, $d(\varphi\lambda) = d\varphi\,\lambda - \varphi\wedge d\lambda$ and $\overline{\varphi\wedge\psi} = -\bar{\psi}\wedge\bar{\varphi}$ for any quaternion valued 1-forms φ and ψ and function $\lambda : M \to I\!H$.

If $s : M \to I\!R^6_1$ is a map into the vector space of quaternionic hermitian forms, then its derivative, $ds : TM \to I\!R^6_1$ is a 1-form with values in the quaternionic hermitian forms. If $Fs \equiv const$, this derivative can be expressed in terms of the connection form Φ of F: since $d(Fs) = 0$,

$$F\,ds = -F[s(.,\Phi) + s(\Phi,.)] \simeq -[Fs\cdot\Phi + \Phi^*\cdot Fs] \tag{6}$$

when using the matrix representation for quaternionic hermitian forms.

5. Curved flats and Isothermic surfaces

The concept of curved flats in symmetric spaces was first introduced by D. Ferus and F. Pedit [9] as a natural generalization of developable surfaces: a curved flat is an envelope of a congruence of flats in a symmetric space or, more technical, a submanifold of a symmetric space (with semisimple isometry group) whose tangent spaces are maximal abelian subalgebras in the tangent spaces of that symmetric space. In [6] it was then applied to the geometry of isothermic surfaces in 3-space. To demonstrate our quaternionic calculus at work, we are going to discuss curved flats in the symmetric space $\mathcal{P}$ of point pairs in $\mathbb{H}P^1$. As in the codimension 1 case, these will turn out to be Darboux pairs of isothermic surfaces in 4-space: given a point pair map $(f,\hat{f}) : M \to \mathcal{P}$, we choose a framing $F : M \to Sl(2,\mathbb{H})$ and write its connection form $\Phi = \Phi_{\mathfrak{k}} + \Phi_{\mathfrak{p}} : TM \to \mathfrak{sl}(2,\mathbb{H}) = \mathfrak{k} \oplus \mathfrak{p}$. Then[10],

Definition. *A map $(f,\hat{f}) : M \to \mathcal{P}$ into the symmetric space of point pairs is called a curved flat if $\Phi_{\mathfrak{p}} \wedge \Phi_{\mathfrak{p}} = 0$.*

Note, that the defining equation is invariant under gauge transformations (3) of F, i.e. does not depend on a choice of homogeneous coordinates. Thus, the notion of a curved flat is a well defined notion for a point pair map $(f,\hat{f}) : M \to \mathcal{P}$.

In order to understand the geometry of a curved flat $(f,\hat{f}) : M^2 \to \mathcal{P}$ in the symmetric space of point pairs we will first express its connection form in a simpler form, and then interpret it geometrically in a second step[11]. We start with an $Sl(2,\mathbb{H})$-framing $F : M^2 \to Sl(2,\mathbb{H})$ and write its connection form

$$\Phi = \begin{pmatrix} \varphi_1 + \varphi_2 j & \hat{\psi}_1 + \hat{\psi}\, j \\ \psi_1 + \psi\, j & \hat{\varphi}_1 + \hat{\varphi}_2 j \end{pmatrix}$$

in terms of complex valued 1-forms. Using a rescaling $(f,\hat{f}) \mapsto (f\lambda, \hat{f}\hat{\lambda})$ we can achieve $\psi_1 = 0$; then, the curved flat equations read (we assume $\psi \neq 0$) $\hat{\psi}_1 = 0$ and $\hat{\psi} \wedge \bar{\psi} = 0$. A rescaling $(f,\hat{f}) \mapsto (f\bar{\lambda}, \hat{f}\lambda^{-1})$ with a complex valued function λ results in $(\psi,\hat{\psi}) \mapsto (\lambda^2\psi, \bar{\lambda}^{-2}\hat{\psi})$; as any 1-form on M^2 has an integrating factor, we may assume $d\psi = 0$, i.e. $\psi = dw$. Since $\hat{\psi} \wedge \bar{\psi} = 0$, $\hat{\psi} = \bar{a}^4 d\bar{w}$ with a suitable function $a : M \to \mathbb{C}$. From the Codazzi equations, $da \wedge dw = 0$ — thus, by a holomorphic change $z_w = a^2$ of coordinates, $\psi = a^{-2}dz$ and $\hat{\psi} = \bar{a}^2 d\bar{z}$, or, after rescaling again with $\lambda = a$, $\psi = dz$ and $\hat{\psi} = d\bar{z}$. Now, the Codazzi equations also yield $\hat{\varphi}_2 \wedge dz = \bar{\varphi}_2 \wedge d\bar{z}$ and $\hat{\varphi}_2 \wedge d\bar{z} = \bar{\varphi}_2 \wedge dz$. Thus, $\varphi_2 = q_1 dz - \bar{q}_2 d\bar{z}$ and $\hat{\varphi}_2 = -\bar{q}_1 dz + q_2 d\bar{z}$ with suitable functions $q_1, q_2 : M \to \mathbb{C}$. This way, $\varphi_2 \wedge \bar{\varphi}_2 = \hat{\varphi}_2 \wedge \bar{\hat{\varphi}}_2$ such that $d\varphi_1 = d\hat{\varphi}_1$ from the Gauss-Ricci equations. With the ansatz $\hat{\varphi}_1 - \varphi_1 = 2a$, we find that a rescaling $(f,\hat{f}) \mapsto (f\lambda, \hat{f}\lambda^{-1})$ with $\lambda = e^a$ yields $\varphi_1 = \hat{\varphi}_1$. At the same time, $(\psi,\hat{\psi}) \mapsto (e^u\psi, e^{-u}\hat{\psi})$ with $u = a + \bar{a}$. So, we end up with a connection form

$$\Phi = \begin{pmatrix} i\eta + (q_1 dz - \bar{q}_2 d\bar{z})j & e^{-u} d\bar{z}\, j \\ e^u dz\, j & i\eta + (-\bar{q}_1 dz + q_2 d\bar{z})j \end{pmatrix} \tag{7}$$

[10] For simplicity of notation, we reduce the definition to the case under investigation.

[11] Note that, from this point on, we will restrict to local geometry: as Darboux pairs of isothermic surfaces generally only exist locally so do curved flats in the space of point pairs. Also, some of the presented arguments require the dondegeneracy of the curvature line net of the surfaces.

where $u : M \to \mathbb{R}$, $q_1, q_2 : M \to \mathbb{C}$ and $\eta : TM \to \mathbb{R}$ is a real valued 1-form — remember that we have chosen an $Sl(2, \mathbb{H})$-framing from the beginning.

In order to interpret this connection form geometrically, we first note that all sphere congruences

$$s_c := F^{-1} \begin{pmatrix} 0 & c \\ \bar{c} & 0 \end{pmatrix} : M \to S^5_1$$

with $c = e^{i\vartheta}$ are enveloped by the two maps f and $\hat{f}$:

$$F ds_c = - \begin{pmatrix} 0 & 2[-\mathrm{Re}(\bar{c}q_1)dz + \mathrm{Re}(cq_2)d\bar{z}]j \\ 2[\mathrm{Re}(\bar{c}q_1)dz - \mathrm{Re}(cq_2)d\bar{z}]j & 0 \end{pmatrix}$$

Thus, in the $\mathbb{R}^6_1$-model of Möbius geometry, the s_c can be viewed as common normal fields of f and $\hat{f}$. Using the identification (1) of points in $\mathbb{H}P^1$ and isotropic lines in $\mathbb{R}^6_1$, we obtain

$$df = F^{-1} \begin{pmatrix} 0 & e^u dz\, j \\ -e^u dz\, j & 0 \end{pmatrix} \quad \text{and} \quad d\hat{f} = F^{-1} \begin{pmatrix} 0 & -e^{-u} d\bar{z}\, j \\ e^{-u} d\bar{z}\, j & 0 \end{pmatrix}$$

as the derivatives (6) of f and $\hat{f}$. Calculating the induced metrics

$$\langle df, df \rangle = e^{2u}|dz|^2 \quad \text{and} \quad \langle d\hat{f}, d\hat{f} \rangle = e^{-2u}|dz|^2$$

of f and $\hat{f}$, and their second fundamental forms with respect to s_c,

$$\begin{array}{rcl} -\langle df, ds_c \rangle & = & e^u[-2\mathrm{Re}(\bar{c}q_1)|dz|^2 + \mathrm{Re}(cq_2)(dz^2 + d\bar{z}^2)], \\ -\langle d\hat{f}, ds_c \rangle & = & e^{-u}[-2\mathrm{Re}(cq_2)|dz|^2 + \mathrm{Re}(\bar{c}q_1)(dz^2 + d\bar{z}^2)], \end{array}$$

we see that f and $\hat{f}$ have well defined principal curvature directions (independent of the normal direction s_c) which do correspond on both surfaces ($\{s_c \,|\, c \in S^1\}$ is a "Ribaucour sphere pencil"), and that f and $\hat{f}$ induce conformally equivalent metrics on M. Moreover, $z : M \to \mathbb{C}$ are conformal curvature line coordinates on both surfaces, i.e. both surfaces are isothermic. Consequently, $(f, \hat{f}) : M \to \mathcal{P}$ is a "Darboux pair" of isothermic surfaces in 4-space[12)]:

DEFINITION. *Two surfaces are said to form a Darboux pair if they envelope a (non-trivial) congruence of 2-spheres (two orthogonal congruences of 3-spheres in 4-space) such that the curvature lines on both surfaces correspond and the induced metrics in corresponding points are conformally equivalent.*

Conversely, if $(f, \hat{f}) : M \to \mathcal{P}$ envelope two congruences of orthogonal spheres, say $s_1, s_i : M \to S^5_1$, then the connection form

$$\Phi = \begin{pmatrix} \varphi_1 + \varphi_2 j & \hat{\psi}\, j \\ \psi\, j & \hat{\varphi}_1 + \hat{\varphi}_2 j \end{pmatrix}$$

12) This geometric description of Darboux pairs of isothermic surfaces can obviously be used to define isothermic surfaces and Darboux pairs of any codimension — as the one below for Christoffel pairs can (cf.[13]). Note, that the flatness of the normal bundle of a surface — which is necessary to make sense of the notion of curvature lines — is a conformal notion, i.e. it is invariant under conformal changes of the ambient space's metric.

with complex 1-forms $\psi, \hat\psi : TM \to \mathbb{C}$. Assuming the curvature lines of f and $\hat f$ to correspond, and their induced metrics to be conformally equivalent, we can introduce common curvature line coordinates: $\psi = e^u\omega$ and $\hat\psi = e^{-u}\omega$, or $\hat\psi = e^{-u}\bar\omega$. In both cases, from the Gauss-Ricci equations $\mathrm{Re}[d(\varphi_1 - \hat\varphi_1)] = 0$, so that after a suitable real rescaling of f and $\hat f$, $\mathrm{Re}(\varphi_1 - \hat\varphi_1) = 0$. Then, in the first case, the Codazzi equations imply $u \equiv const$: the sphere congruences enveloped by f and $\hat f$ lie in a fixed linear complex, consequently f and $\hat f$ are congruent in some space of constant curvature (cf.[3], [6]) — and are not considered to form a Darboux pair. In the other case, the Codazzi equations yield $d\omega = 0$ — we have *conformal* curvature line parameters, i.e. f and $\hat f$ are isothermic; we could also have concluded this from the fact that f and $\hat f$ obviously form a curved flat:

Theorem. *A surface pair $(f, \hat f) : M^2 \to \mathcal{P}$ is a curved flat if and only if f and $\hat f$ form a Darboux pair. Two surfaces forming a Darboux pair are isothermic.*

The $\mathfrak{k}$-part — see (2) — of the Maurer-Cartan equation of a $Gl(2, \mathbb{H})$-framing reads $0 = d\Phi_{\mathfrak{k}} + \Phi_{\mathfrak{k}} \wedge \Phi_{\mathfrak{k}} + \Phi_{\mathfrak{p}} \wedge \Phi_{\mathfrak{p}}$. Thus, for a curved flat, $\Phi_{\mathfrak{k}} = H^{-1}dH$ with a suitable $H : M \to K$: if λ and $\hat\lambda$ are given by

$$\lambda^{-1}d\lambda = i\eta + (q_1 dz - \bar q_2 d\bar z)j \quad \text{and} \quad \hat\lambda^{-1}d\hat\lambda = i\eta + (-\bar q_1 dz + q_2 d\bar z)j$$

then a gauge transformation $(f, \hat f) \mapsto (f\lambda^{-1}, \hat f\hat\lambda^{-1})$ of our previous framing with connection form (7) leaves us with

$$\Phi = \begin{pmatrix} 0 & \lambda(e^{-u}d\bar z\, j)\hat\lambda^{-1} \\ \hat\lambda(e^u dz\, j)\lambda^{-1} & 0 \end{pmatrix} =: \begin{pmatrix} 0 & \hat\omega \\ \omega & 0 \end{pmatrix}.$$

The Codazzi equations for this new framing simply read $d\omega = d\hat\omega = 0$ showing that $\bar\omega = df_0$ and $\hat\omega = d\hat f_0$ with suitable maps $f_0, \hat f_0 : M \to \mathbb{H}$. Here, we identify the two copies of the quaternions sitting in $\mathfrak{p} = \mathbb{H} \oplus \mathbb{H}$ as the eigenspaces of $\mathrm{ad}_C : \mathfrak{p} \to \mathfrak{p}$, $C = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, by means of the real endomorphism $X \mapsto X^*$ of $\mathfrak{p}$. Note, that since the 1-forms $\lambda^{-1}d\lambda, \hat\lambda^{-1}d\hat\lambda : TM \to \mathrm{Im}\mathbb{H}$ take values in the imaginary quaternions, $|\lambda| = |\hat\lambda| \equiv 1$. Consequently, the induced metrics of $f_0 : M \to \mathbb{H}$ and $\hat f_0 : M \to \mathbb{H}$, $\mathbb{H} \cong \mathbb{R}^4$ considered as a Euclidean space, are

$$df_0 \cdot df_0 = e^{2u}|dz|^2 \quad \text{and} \quad d\hat f_0 \cdot d\hat f_0 = e^{-2u}|dz|^2.$$

Moreover, with the common unit normal fields $n_c = -\lambda c\hat\lambda^{-1}$ of f_0 and $\hat f_0$, where $c = e^{i\vartheta}$, their second fundamental forms become

$$\begin{array}{rcl} -df_0 \cdot dn_c & = & e^u[-2\mathrm{Re}(cq_1)|dz|^2 + \mathrm{Re}(\bar c q_2)(dz^2 + d\bar z^2)], \\ -d\hat f_0 \cdot d\hat n_c & = & e^{-u}[-2\mathrm{Re}(\bar c q_2)|dz|^2 + \mathrm{Re}(c q_1)(dz^2 + d\bar z^2)]. \end{array} \tag{8}$$

Thus, f_0 and $\hat f_0$ are two isothermic surfaces that carry common curvature line coordinates — and, $\hat f_0$ and $\bar f_0$ have parallel tangent planes. Hence, we define[13]:

[13] If $f_0, \hat f_0 : M^2 \to \mathrm{Im}\mathbb{H}$, this definition yields the classical notion of a Christoffel pair (cf.[6]).

Definition. *Two (non homothetic) surfaces* $f_0, \hat{f}_0 : M^2 \to I\!H$ *with parallel tangent planes in corresponding points are said to form a Christoffel pair if the curvature lines on both surfaces correspond and the induced metrics are conformally equivalent.*

Conversely, if two surfaces $f_0, \hat{f}_0 : M^2 \to I\!H$ carry conformally equivalent metrics and have parallel tangent planes in corresponding points $f_0(p)$ and $\hat{f}_0(p)$ then[14], $df_0 = \lambda e^u \psi j \hat{\lambda}^{-1}$ and $d\hat{f}_0 = \pm\lambda e^{-u} \psi j \hat{\lambda}^{-1}$, or $d\hat{f}_0 = \lambda e^{-u} \bar{\psi} j \hat{\lambda}^{-1}$ with a real valued function u, a complex 1-form $\psi : TM \to \mathbb{C}$ and suitable quaternionic functions $\lambda, \hat{\lambda} : M \to I\!H$ — where $|\lambda| = |\hat{\lambda}| \equiv 1$ without loss of generality. In the first case, the integrability conditions yield $0 = du \wedge \psi$ showing that $u \equiv const$. Consequently, $\hat{f}_0$ is homothetic to f_0 — and f_0 and $\hat{f}_0$ are not considered to form a Christoffel pair. In the second case, $d\bar{f}_0 \wedge d\hat{f}_0 = d\hat{f}_0 \wedge d\bar{f}_0 = 0$. Hence, the surface pair $f_0, \hat{f}_0 : M \to I\!H$ gives rise to a curved flat by integrating $\Phi := \begin{pmatrix} 0 & d\hat{f}_0 \\ d\bar{f}_0 & 0 \end{pmatrix}$ — we obtain the following

Theorem. *Two surfaces* $f_0, \hat{f}_0 : M^2 \to I\!H$ *form a Christoffel pair if and only if* $d\bar{f}_0 \wedge d\hat{f}_0 = d\hat{f}_0 \wedge d\bar{f}_0 = 0$. *Two surfaces forming a Christoffel pair are isothermic.*

Curved flats — or, Darboux pairs of isothermic surfaces — naturally arise in 1-parameter families [9]: if $\Phi = \Phi_{\mathfrak{k}} + \Phi_{\mathfrak{p}}$ denotes one of the connection forms associated to a curved flat $(f, \hat{f}) : M^2 \to \mathcal{P}$, then, with a real parameter $\varrho \in I\!R$, all the connection forms

$$\Phi_\varrho := \Phi_{\mathfrak{k}} + \varrho^2 \Phi_{\mathfrak{p}} : TM^2 \to \mathfrak{sl}(2, I\!H) = \mathfrak{k} \oplus \mathfrak{p} \tag{9}$$

are integrable and give rise to curved flats $(f_\varrho, \hat{f}_\varrho) : M^2 \to \mathcal{P}$; in fact, if the connection forms (9) are integrable for more than one value of ϱ^2, then the associated point pair maps are necessarily curved flats. From (3), we learn that this 1-parameter family of curved flats does not depend on the framing chosen to describe the curved flat $(f, \hat{f})$. Moreover, sending the parameter $\varrho \to 0$, and rescaling $(f_\varrho, \hat{f}_\varrho) \mapsto (\varrho^{-1} f_\varrho, \varrho \hat{f}_\varrho)$ or $(f_\varrho, \hat{f}_\varrho) \mapsto (\varrho f_\varrho, \varrho^{-1} \hat{f}_\varrho)$ at the same time, provides us with

$$(f_{\varrho=0}, \hat{f}_{\varrho=0}) = \begin{pmatrix} 1 & 0 \\ \bar{f}_0 & 1 \end{pmatrix} \quad \text{or} \quad (f_{\varrho=0}, \hat{f}_{\varrho=0}) = \begin{pmatrix} 1 & \hat{f}_0 \\ 0 & 1 \end{pmatrix}.$$

Hence, we may think of the Christoffel pair $(f_0, \hat{f}_0)$ — that is, as before, associated to a 1-parameter family of curved flats by integrating

$$\Phi_\varrho = \begin{pmatrix} 0 & \varrho^2 d\hat{f}_0 \\ \varrho^2 d\bar{f}_0 & 0 \end{pmatrix}$$

— as a limiting case for the Darboux pairs $(f_\varrho, \hat{f}_\varrho)$. Comparison with (3) shows that the spectral parameter ϱ corresponds to the scaling ambiguity of the members of a Christoffel pair: one of the surfaces of a Christoffel pair is determined by the other only up to a homothety (and translation).

We will use those facts to discuss perturbation methods (cf.[16]) for the construction of constant mean curvature surfaces and, in particular, for Bryant's Weierstrass type representation [4] for

[14] If p is not an umbilic for either surface, it *follows* that the principal curvature directions of both surfaces correspond. In case one of the surfaces is totally umbilic we need also to assume that the curvature lines on both surfaces coincide — otherwise we might find two associated minimal surfaces.

6. CONSTANT MEAN CURVATURE SURFACES

in hyperbolic space forms. We restrict our attention to codimension 1 by assuming that our surfaces lie in a fixed conformal 3-sphere, say s_1. Thus the connection form (7) of a Darboux pair $(f,\hat{f}) : M^2 \to \mathbb{H}P^1$ takes the form

$$\Phi = \begin{pmatrix} i[\eta + \frac{1}{2}(e^u H dz - e^{-u}\hat{H}d\bar{z})j] & e^{-u}d\bar{z}\,j \\ e^u dz\,j & i[\eta + \frac{1}{2}(e^u H dz - e^{-u}\hat{H}d\bar{z})j] \end{pmatrix} \tag{10}$$

where the (real) functions $H, \hat{H}$ can be interpreted as the mean curvature functions of the members f_0 and $\hat{f}_0$ of the limiting Christoffel pair: from (10) we see that a rescaling $(f,\hat{f}) \mapsto (f\lambda, \hat{f}\lambda)$ will provide us with $\Phi_{\mathfrak{k}} = 0$, such that $df_0, d\hat{f}_0 : TM \to \mathrm{Im}\mathbb{H}$. The second fundamental forms (8) with respect to the remaining common normal field $n_i = -\lambda i \lambda^{-1} = -\hat{n}_i$ become

$$\begin{aligned} -df_0 \cdot dn_i &= He^{2u}|dz|^2 - \tfrac{1}{2}\hat{H}(dz^2 + d\bar{z}^2)], \\ -d\hat{f}_0 \cdot d\hat{n}_i &= \hat{H}e^{-2u}|dz|^2 - \tfrac{1}{2}H(dz^2 + d\bar{z}^2)]. \end{aligned}$$

The Codazzi equations (5) yield $\eta = \frac{i}{2}(-u_z dz + u_{\bar{z}}d\bar{z})$ and from (4) we recover the classical Gauss equation $0 = u_{z\bar{z}} + \frac{1}{4}(H^2e^{2u} - \hat{H}^2e^{-2u})$ holding for both surfaces f_0 and $\hat{f}_0$, and the classical Codazzi equations $dH \wedge e^u dz = d\hat{H} \wedge e^{-u}d\bar{z}$. Hence, $H \equiv const$ if and only if $\hat{H} \equiv const$, reflecting the fact that a pair of parallel constant mean curvature surfaces, or a minimal surface and its Gauss map form Christoffel pairs (cf.[11]).

Calculating the derivative of the sphere congruence s_i enveloped by the two surfaces f and $\hat{f}$ — which form the Darboux pair associated with the Christoffel pair $(f_0, \hat{f}_0)$ — we find

$$F ds_i = \begin{pmatrix} 0 & (He^u dz - \hat{H}e^{-u}d\bar{z})j \\ (-He^u dz + \hat{H}e^{-u}d\bar{z})j & 0 \end{pmatrix} = H \cdot F\,df + \hat{H} \cdot F\,d\hat{f}.$$

Hence, the vector $N := s_i - Hf - \hat{H}\hat{f}$ is constant as soon as one of the mean curvatures, H or $\hat{H}$, is. In order to interpret this fact geometrically, we have to distinguish two cases:

If $H\hat{H} \neq 0$, i.e. $(f_0, \hat{f}_0)$ is equivalent to a pair of parallel constant mean curvature surfaces, then $\langle N, \frac{2}{\hat{H}}f\rangle \equiv 1$ and $\langle N, \frac{2}{H}\hat{f}\rangle \equiv 1$ — and consequently (cf.[3]), the two surfaces $\frac{1}{\hat{H}}f, \frac{1}{H}\hat{f} : M^2 \to s_1 \simeq S^3 \subset \mathbb{H}P^1$ can be interpreted as surfaces in the space $M^3_N := \{y \in \mathbb{R}^6_1 \,|\, \langle N, y\rangle = 1, \langle s_1, y\rangle = 0, \langle y, y\rangle = 0\}$ of constant sectional curvature $\kappa = -\langle N, N\rangle = -(1 - H\hat{H})$. Their induced metrics are

$$\langle d(\tfrac{2}{\hat{H}}f), d(\tfrac{2}{\hat{H}}f)\rangle = \tfrac{4}{\hat{H}^2}e^{2u}|dz|^2 \quad \text{and} \quad \langle d(\tfrac{2}{H}\hat{f}), d(\tfrac{2}{H}\hat{f})\rangle = \tfrac{4}{H^2}e^{-2u}|dz|^2$$

while, with the unit normal fields $t = s_i - \frac{2}{\hat{H}}f$ and $\hat{t} = s_i - \frac{2}{H}\hat{f}$ in that space M^3_N, their second fundamental forms become

$$\begin{aligned} -\langle d(\tfrac{2}{\hat{H}}f), dt\rangle &= \tfrac{4}{\hat{H}^2}e^{2u}(1 - \tfrac{1}{2}H\hat{H})\,|dz|^2 + (dz^2 + d\bar{z}^2) \\ -\langle d(\tfrac{2}{H}\hat{f}), d\hat{t}\rangle &= \tfrac{4}{H^2}e^{-2u}(1 - \tfrac{1}{2}H\hat{H})\,|dz|^2 + (dz^2 + d\bar{z}^2) \end{aligned}$$

— showing that both surfaces have the same constant mean curvature $1 - \frac{1}{2}H\hat{H}$. As a special case, $H = 1$ and $\hat{H} = 2$, this provides the well known relation between constant mean curvature surfaces in Euclidean space $\mathbb{R}^3$ and minimal surfaces in the 3-sphere S^3.

If $H\hat{H} = 0$, one of the surfaces f_0 or $\hat{f}_0$ is a minimal surface, say $\hat{H} = 0$, while the other is homothetic to its Gauss map, say $n = H\, f_0$. Now, the surface $\frac{2}{H}\hat{f} : M^2 \to M_N^3$ lies in hyperbolic space, $\kappa = -1$, while f is the hyperbolic Gauss map (cf.[4]) of $\frac{2}{H}\hat{f}$ since $\langle N, f\rangle \equiv 0$, i.e. f takes values in the infinity boundary $N \in S_1^5$ of M_N^3. As before, the mean curvature of $\frac{2}{H}\hat{f} : M^2 \to M_N^3$ is easily calculated to be constant $= 1$. This is how Bryant's Weierstrass type representation [4] for surfaces of constant mean curvature 1 in hyperbolic 3-space H^3 can be obtained in this context: we write the differential $d\hat{f}_0 = \frac{1}{2}(i + gj)\bar{\omega}j(i + gj)$ of a minimal immersion $\hat{f}_0 : M^2 \to \mathbb{R}^3$ (and its Christoffel transform, its Gauss map $f_0 = (i + gj)i(i + gj)^{-1} : M^2 \to S^2$) in terms of a holomorphic 1-form $\omega : TM^2 \to \mathbb{C}$ and the (meromorphic) stereographic projection $g : M \to \mathbb{C}$ of its Gauss map. Then, the constant mean curvature surface $\hat{f} : M^2 \to H^3$ (and its hyperbolic Gauss map $f : M^2 \to N \simeq S^2$) are obtained by integrating the connection form[15)]

$$\Phi = \begin{pmatrix} 0 & \frac{1}{2}(i + gj)\bar{\omega}j(i + gj) \\ -2(i + gj)^{-1}dg\, j(i + gj)^{-1} & 0 \end{pmatrix}, \tag{11}$$

to the framing $(f, \hat{f}) \simeq F : M^2 \to Gl(2, \mathbb{H})$ where $dF = F\Phi$ — thus (locally) characterizing Bryant's Weierstrass type representation of surfaces of constant mean curvature 1 in hyperbolic space as Bianchi's T-transform [2] of minimal surfaces in Euclidean space. In fact, introducing the spectral parameter (9), surfaces of constant mean curvature c in hyperbolic space forms of curvature $\kappa = -c^2$ arise by "perturbation" of minimal surfaces in Euclidean 3-space (cf.[16]).

Parametrizing a minimal surface patch $\hat{f}_0$ in terms of curvature line parameters, $z = x + iy$, the above representation of $\hat{f}_0$ becomes the classical Enneper-Weierstrass representation, i.e. $\omega = \frac{dz}{g'}$. Performing a Möbius transformation on the Gauss map g (resp. f_0 — its Christoffel transform) and integrating the Enneper-Weierstrass representation again (i.e. taking the Christoffel transform of the Möbius transformed Gauss map) yields the classical Goursat transformation of the minimal surface patch. But, a closer look at the connection form (11) suggests that the Enneper-Weierstrass representation itself can be interpreted as a Goursat type transformation of a planar patch: considering $gj, \int\bar{\omega}j : M^2 \to \mathbb{C}j$ as a (highly degenerate) Christoffel pair, the corresponding minimal surface $\hat{f}_0$ is obtained as a Christoffel transformation of $f_0 = \frac{1}{1+|g|^2}[(1 - |g|^2)i + 2gj]$, the stereographic projection of gj ("the" Christoffel transform of $\int\bar{\omega}j$) into S^2. This Goursat type transformation can (obviously) be generalized to arbitrary Christoffel pairs of isothermic surfaces: if $f_0, \hat{f}_0 : M^2 \to \mathbb{H}$ form a Christoffel pair, then, for any (constant) $a \in \mathbb{H}$, the quaternionic 1-forms

15) With the ansatz $F = \begin{pmatrix} 2(\overline{x_{21}g + x_{22}})(i + gj)^{-1} & j(x_{21}i - x_{22}j) \\ 2j(\overline{x_{11}g + x_{12}})(i + gj)^{-1} & -(x_{11}i - x_{12}j) \end{pmatrix}$, the common form of Bryant's representation is obtained as $xx^* : M^2 \to H^3 \subset \{y \in Gl(2, \mathbb{C}) \,|\, y = y^*\} \cong \mathbb{R}_1^4$ where the scalar product on H^3 is induced by the Lorentz scalar product $|y|^2 = -\det(y)$ on $\mathbb{R}_1^4$.

$(a+\bar{f}_0)^{-1}d\bar{f}_0(a+\bar{f}_0)^{-1}$ and $(a+\bar{f}_0)d\hat{f}_0(a+\bar{f}_0)$ are closed — and consequently give rise to a new Christoffel pair.

Acknowledgements: I would like to thank the many people who contributed to this paper by questions and discussions. In particular, I would like to thank the members of the GANG, Franz Pedit, my host at GANG, George Kamberov, Catherine McCune, Ian McIntosh, Lorenzo Nicolodi and Ulrich Pinkall. Also, I would like to thank the Referee for his comments and questions concerning the first version of the paper.

References

1. H. Aslaksen: *Quaternionic Determinants*; Math. Intell. 18.3 (1996) 57-65
2. L. Bianchi: *Complementi alle Ricerche sulle Superficie Isoterme*; Ann. Mat. pura appl. 12 (1905) 19–54
3. W. Blaschke: *Vorlesungen über Differentialgeometrie III*; Springer, Berlin 1929
4. R. Bryant: *Surfaces of mean curvature one in hyperbolic space*; Astérisque 154-155 (1987) 321-347
5. F. Burstall: personel communication, 1997
6. F. Burstall, U. Hertrich-Jeromin, F. Pedit, U. Pinkall: *Isothermic surfaces and Curved flats*; Math. Z. 225 (1997) 199-209
7. E. Christoffel: *Ueber einige allgemeine Eigenschaften der Minimumsflächen*; Crelle's J. 67 (1867) 218–228
8. G. Darboux: *Sur les surfaces isothermiques*; Comptes Rendus 122 (1899) 1299–1305, 1483–1487, 1538
9. D. Ferus, F. Pedit: *Curved flats in Symmetric spaces*; Manuscripta Math. 91 (1996) 445-454
10. U. Hertrich-Jeromin, T. Hoffmann, U. Pinkall: *A discrete version of the Darboux transform for isothermic surfaces*; to appear in A. Bobenko, R. Seiler, *Discrete integrable Geometry and Physics*, Oxford Univ. Press, Oxford 1997
11. U. Hertrich-Jeromin, F. Pedit: *Remarks on the Darboux transform of isothermic surfaces*; Doc. Math. J. DMV 2 (1997) 313–333
12. G. Kamberov: *Holomorphic quadratic differentials and Conformal immersions*; Preprint 1997
13. B. Palmer: *Isothermic surfaces and the Gauss map*; Proc. Amer. Math. Soc. 104 (1988) 876-884
14. E. Study: *Ein Seitenstück zur Theorie der linearen Transformationen einer komplexen Veränderlichen, Teile I-IV*; Math. Z. 18 (1923) 55-86, 201-229 and 21 (1924) 45-71, 174-194
15. J. Wilker: *The Quaternion formalism for Möbius groups in four or fewer dimensions*; Lin. Alg. Appl. 190 (1993) 99-136
16. M. Umehara, K. Yamada: *A parametrization of the Weierstrass formulae and perturbation of complete minimal surfaces in $\mathbb{R}^3$ into the hyperbolic 3-space*; J. reine angew. Math. 432 (1992) 93-116

Udo Hertrich-Jeromin
Dept. Math. & Stat., GANG
University of Massachusetts
Amherst, MA 01003
jeromin@math.umass.edu

On the Cuspidal Divisor Class Group of a Drinfeld Modular Curve

Ernst-Ulrich Gekeler

Received: September 11, 1997

Communicated by Peter Schneider

Abstract. The theory of theta functions for arithmetic groups Γ that act on the Drinfeld upper half-plane is extended to allow degenerate parameters. This is used to investigate the cuspidal divisor class groups of Drinfeld modular curves. These groups are finite for congruence subgroups Γ and may be described through the corresponding quotients of the Bruhat-Tits tree by Γ. The description given is fairly explicit, notably in the most important special case of Hecke congruence subgroups Γ over a polynomial ring.

1991 Mathematics Subject Classification: 11G09, 11G18, 11F11, 11F12

Keywords: Drinfeld modular curves, theta functions, cuspidal divisor class groups

Introduction.

Drinfeld modular curves are the substitutes in positive characteristics of classical modular curves. Like these, they have a rich structure where various mathematical disciplines interact: number theory, algebraic geometry, (non-Archimedean) function theory, representation theory and automorphic forms, and others. They encode important pieces of the arithmetic of global function fields, notably those related to two-dimensional Galois representations and elliptic curves, in a way similar to the correspondence ascribed to Shimura, Taniyama and Weil and partially proven by A. Wiles.

By their very construction, these curves come equipped with a uniformization through the Drinfeld upper half-plane Ω, a one-dimensional rigid analytic symmetric space. Hence many questions about such a curve M_Γ may be attacked by function theoretic means, through the construction and investigation of analytic functions on Ω (analogues of elliptic modular forms, or of theta functions) that satisfy functional equations under Γ, the group that uniformizes $M_\Gamma = \Gamma \setminus \Omega$.

Leaving aside Tate's elliptic curves, the first appearance of non-Archimedean uniformized curves is in work of Mumford [16] and of Manin-Drinfeld [14], where the acting group Γ is a Schottky group, that is, a finitely generated free group consisting of hyperbolic elements. For the corresponding *Mumford curves*, Gerritzen and van der Put in their monograph [11] obtained a very satisfactory description of the minimal model, the Jacobian, the Abel-Jacobi map, ...

A similar program for Drinfeld modular curves was started in [10], whose main results were the construction of the Jacobian J_Γ of M_Γ through non-Archimedean theta functions $\theta_\Gamma(\omega, \eta, z)$ and, as an application, the analytic description of "Weil uniformizations" of elliptic curves over global functions fields. Apart from the fact that a Drinfeld modular curve is defined over a global field (which gives an abundance of arithmetic structure), the crucial difference to Mumford curves is that $M_\Gamma = \Gamma \setminus \Omega$ by construction is an affine curve, and has to be "compactified" to a smooth projective curve $\overline{M}_\Gamma$ by adding a finite number of "cusps" of Γ. Several natural questions (with important arithmetical applications) arise, about the

- structure of the group $\mathcal{C}$ generated in the Jacobian J_Γ by the cusps;
- degeneration of the theta functions $\theta_\Gamma(\omega, \eta, z)$ if the parameters $\omega, \eta \in \Omega$ approach cusps of Γ;
- relationship between $\mathcal{C}$ and the minimal model of $\overline{M}_\Gamma$.

It turns out that these questions have satisfactory answers in terms of the associated *almost finite* graphs $\Gamma \setminus \mathcal{T}$, which can be mechanically calculated from the initial data that define Γ, e.g., from congruence conditions.

In order to give more precise statements, we now introduce some notation.

We start with a function field K in one variable with exact field of constants $\mathbb{F}_q$, the finite field with $q = p^r$ elements. In K, we fix a place "∞", and we let $A \subset K$ be the Dedekind subring of elements regular away from ∞. Then A is a discrete and cocompact subring of the completion K_∞. We finally need C, the completed algebraic closure of K_∞. By an arithmetic subgroup of $\mathrm{GL}(2, K)$, we understand a subgroup commensurable with $\mathrm{GL}(2, A)$. Such a group Γ acts with finite stabilizers on $\Omega = C - K_\infty$, and M_Γ will be the uniquely determined algebraic curve whose space of C-points is given by $\Gamma \setminus \Omega$. The cusps are given as the orbits $\Gamma \setminus \mathbb{P}^1(K)$ on the projective line $\mathbb{P}^1(K)$. It is customary to recall here the obvious analogy of the data K, A, K_∞, C, Ω, $\mathrm{GL}(2, A)$ with $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$, H = complex upper half-plane, $\mathrm{SL}(2, \mathbb{Z})$ (or rather $H^\pm = \mathbb{C} - \mathbb{R}$ and $\mathrm{GL}(2, \mathbb{Z})$), respectively.

In [10], we studied theta functions $\theta_\Gamma(\omega, \eta, z)$, which are defined as certain infinite products depending on parameters $\omega, \eta \in \Omega$. These functions are meromorphic on Ω with zeros (resp. poles) at the orbits of ω (resp. η); they transform according to a character $c(\omega, \eta) : \Gamma \longrightarrow C^*$, have a nice behavior at the boundary $\partial\Omega = \mathbb{P}^1(K)$ of Ω, and give rise to a pairing $\overline{\Gamma} \times \overline{\Gamma} \longrightarrow K_\infty^*$ on the maximal torsion-free Abelian quotient $\overline{\Gamma}$ of Γ. The analytic space Ω has a canonical covering through standard rational subsets of $\mathbb{P}^1(C)$, the nerve of which equals the Bruhat-Tits tree $\mathcal{T}$ of $\mathrm{GL}(2, K_\infty)$. There results a $\mathrm{GL}(2, K_\infty)$-equivariant map $\lambda : \Omega \longrightarrow \mathcal{T}(\mathbb{R})$ that allows to describe many properties of M_Γ and of related objects in terms of the graph $\Gamma \setminus \mathcal{T}$. The main results of the present paper go into this direction. They are:

- Theorem 3.8 and its corollaries, which give the link between theta functions, cuspidal divisors on $\overline{M}_\Gamma$, and harmonic Γ-invariant cochains on $\mathcal{T}$;

- the description, given in sections 4 and 5, of the cuspidal divisor class group $\mathcal{C}(\Gamma)$ of $\overline{M}_\Gamma$ and of the canonical map from $\mathcal{C}(\Gamma)$ to $\Phi_\infty(\Gamma)$ = group of connected components of the Néron model of J_Γ at ∞ (here Γ is assumed to be a congruence subgroup);
- the determination of the subgroup generated by the $\theta_\Gamma(\omega, \eta, z)$ ($\omega, \eta \in \mathbb{P}^1(K)$) in the group of all theta functions for Γ (Thm. 5.4), valid for Hecke congruence subgroups Γ of $\mathrm{GL}(2, A)$, where A is a polynomial ring.

These results depend on an extension of the theory developed in [10] to the case of theta functions $\theta_\Gamma(\omega, \eta, z)$ whose parameters ω, η are allowed to lie in the boundary of Ω. This is carried out in section two: proof of convergence, functional equation, behavior at the boundary. Roughly speaking, theta functions with degenerate parameters behave similar to those with $\omega, \eta \in \Omega$, and analytic dependence on the parameters holds at least for the associated multipliers $c(\omega, \eta)$. That part of the theory, as well as the links (given in section three) with harmonic cochains on $\mathcal{T}$ and cuspidal divisor groups on $\overline{M}_\Gamma$, works in the context of arbitrary groups Γ commensurable with $\mathrm{GL}(2, A)$, and may thus be used also for the study of non-congruence subgroups. From section four on we specialize to congruence subgroups Γ and use the known finiteness of $\mathcal{C}(\Gamma)$ in this case (i.e., the analogue of Manin-Drinfeld's theorem, cf. [2], [5]) to express it through the graph $\Gamma \setminus \mathcal{T}$. $\mathcal{C}(\Gamma)$ agrees (modulo finite groups annihilated by $q^{\deg\,\infty} - 1$) with $\underline{H}/\underline{H}_! \oplus \underline{H}_!^\perp$, where $\underline{H} = \underline{H}(\mathcal{T}, \mathbb{Z})^\Gamma$ is the group of Γ-invariant $\mathbb{Z}$-valued harmonic cochains on $\mathcal{T}$, $\underline{H}_!$ is the subgroup of cochains with compact support mod Γ, and $\underline{H}_!^\perp$ its ortho-complement in $\underline{H}$.

A refinement of the above in the important special case of Hecke congruence subgroups $\Gamma_0(n)$ over $A = \mathbb{F}_q[T]$ is given in section five. Here we use in a crucial way the known results (cf. [9]) about the structure of the graph $\Gamma_0(n) \setminus \mathcal{T}$. We conclude, in section six, with a worked-out example (hopefully instructive), where the canonical map $\mathrm{can}_\infty : \mathcal{C}(\Gamma) \longrightarrow \Phi_\infty(\Gamma)$ fails to be injective or surjective even for a Hecke congruence group Γ with prime conductor. The existence of a non-trivial kernel of can_∞ is reflected in congruence properties of a corresponding "Eisenstein quotient" of J_Γ, an elliptic curve in the example treated.

The notation of the present paper is largely compatible to that of [10], to which it is a sequel. Thus without further explanation, for a group G acting on a set X and $x \in X$, G_x is the stabilizer, Gx the orbit, $G \setminus X$ the set of all orbits, G^{ab} the maximal Abelian quotient of G. We often write gx for $g(x)$, $g \in G$. As far as misconceptions are unlikely, we do not distinguish between matrices in $\mathrm{GL}(2)$ and their classes in $\mathrm{PGL}(2)$, and between varieties over C or K_∞, their associated analytic spaces, and their sets of C-valued points.

1. Background [10].

(1.1) We let K be the function field of a smooth projective geometrically connected curve $\mathfrak{C}$ over $\mathbb{F}_q$ (q = power of the rational prime p) and $\infty \in \mathfrak{C}$ a closed point fixed once for all. Attached to these data, we dispose of

- the subring A of K of functions regular away from ∞;
- the completion K_∞ of K at ∞;
- the completed algebraic closure $C = C_\infty$ of K_∞;

- Drinfeld's upper half-plane $\Omega = C - K_\infty$, on which $\mathrm{GL}(2, K_\infty)$ acts through $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) z = \frac{az+b}{cz+d}$;
- the Bruhat-Tits tree $\mathcal{T}$ of $\mathrm{GL}(2, K_\infty)$.

Recall that $\mathcal{T}$ is a $(q_\infty + 1)$-regular tree ($q_\infty = q^{\deg \infty}$ = size of residue class field $\mathbb{F}_q(\infty)$) provided with a $\mathrm{GL}(2, K_\infty)$-action and an equivariant map λ from Ω to the real points $\mathcal{T}(\mathbb{R})$ of $\mathcal{T}$.

The group $\mathrm{GL}(2, K)$ acts from the right on the space K^2 of row vectors. For an A-lattice (= projective A-submodule of rank two) $Y \hookrightarrow K^2$, we let $\mathrm{GL}(Y) = \{\gamma \in \mathrm{GL}(2, K) \mid Y\gamma = Y\}$.

(1.2) An *arithmetic subgroup* Γ of $\mathrm{GL}(2, K)$ is a subgroup commensurable with some $\mathrm{GL}(Y)$, i.e., $\Gamma \cap \mathrm{GL}(Y)$ has finite index in both Γ and $\mathrm{GL}(Y)$, and which acts without inversion on $\mathcal{T}$. A *congruence subgroup* is some Γ that satisfies $\mathrm{GL}(Y, \mathfrak{n}) \subset \Gamma \subset \mathrm{GL}(Y)$, where $0 \neq \mathfrak{n} \subset A$ is an ideal and $\mathrm{GL}(Y, \mathfrak{n})$ is the kernel of the reduction map $\mathrm{GL}(Y) \longrightarrow \mathrm{GL}(Y/\mathfrak{n}Y)$. According to [20] II Thm. 12, there are "many" subgroups of finite index of $\mathrm{GL}(Y)$ that are not congruence subgroups, although it is not easy to display examples.

Now fix some arithmetic subgroup Γ as above. The following facts, in the case of congruence subgroups, are proved and/or described in more detail in [10] I - III; their generalization to arbitrary arithmetic subgroups is obvious .

(1.2.1) Γ acts with finite stabilizers on Ω and $\mathcal{T}$. Hence e.g. the quotient $\Gamma \setminus \Omega$ may be defined as an analytic space.

(1.2.2) Γ has finite covolume in $\mathrm{GL}(2, K_\infty)$ modulo its center.

(1.2.3) The quotient $\Gamma \setminus \mathcal{T}$ is (in an essentially unique fashion, *loc. cit.*) the union of a finite graph and a finite number of half-lines $\bullet - - - \bullet - - - \bullet - - - \bullet \cdots$, the *ends* of $\Gamma \setminus \mathcal{T}$.

(1.2.4) There exists a smooth connected affine algebraic curve M_Γ / C (which may even be defined over a finite field extension $K' \subset K_\infty$ of K) whose set $M_\Gamma(C)$ of C-points agrees with $\Gamma \setminus \Omega$ as an analytic space. The M_Γ or their canonical smooth compactifications $\overline{M}_\Gamma$ are what we here call Drinfeld modular curves.

(1.2.5) There are canonical bijections between the sets of

(a) ends of $\Gamma \setminus \mathcal{T}$,

(b) cusps $\overline{M}_\Gamma(C) - M_\Gamma(C)$ of $\overline{M}_\Gamma$, and

(c) orbits $\Gamma \setminus \mathbb{P}^1(K)$ on the projective line $\mathbb{P}^1(K)$.

In the sequel, we will not distinguish between (a), (b), (c) and label it by *cusp*(Γ). Its cardinality is denoted by $c = c(\Gamma)$.

(1.2.6) The genus $g = g(\Gamma)$ of $\overline{M}_\Gamma$ agrees with the number of $\dim_{\mathbb{Q}} H_1(\Gamma \setminus \mathcal{T}, \mathbb{Q})$ of independent cycles of the graph $\Gamma \setminus \mathcal{T}$, which in turn equals the rank $\mathrm{rk}(\Gamma^{ab})$ of the factor commutator group Γ^{ab} of Γ.

Let $\overline{\Gamma} = \Gamma^{ab}/\mathrm{tor}(\Gamma^{ab}) \cong \mathbb{Z}^{g(\Gamma)}$ and Γ_f be the subgroup of Γ generated by the elements of finite order. It follows from [20] I Thm. 13, Cor. 1 that

(1.2.7) (i) Γ/Γ_f is free in g generators,

(ii) $\mathrm{tor}(\Gamma^{ab})$ is generated by the image of Γ_f in Γ^{ab}, and

(iii) the canonical map $\overline{\Gamma} \longrightarrow (\Gamma/\Gamma_f)^{ab}$ is an isomorphism.

(1.3) Let $X(\mathcal{T})$ and $Y(\mathcal{T})$ be the sets of vertices, of oriented edges of $\mathcal{T}$, respectively. As in [10], $\underline{H}(\mathcal{T},\mathbb{Z})$ is the right $\mathrm{GL}(2,K_\infty)$-module of $\mathbb{Z}$-valued harmonic cochains in $\mathcal{T}$, i.e., of maps $\varphi : Y(\mathcal{T}) \longrightarrow \mathbb{Z}$ that satisfy $\varphi(\overline{e}) = -\varphi(e)$ ($\overline{e} = e$ oriented inversely) and

$$\sum_{e\in Y(\mathcal{T}) \text{ with origin } v} \varphi(e) = 0 \qquad (v \in X(\mathcal{T})). \tag{1.3.1}$$

Further, $\underline{H}(\mathcal{T},\mathbb{Z})^\Gamma$ denotes the Γ-invariants in $\underline{H}(\mathcal{T},\mathbb{Z})$ and $\underline{H}_!(\mathcal{T},\mathbb{Z})^\Gamma \subset \underline{H}(\mathcal{T},\mathbb{Z})^\Gamma$ the subgroup of those φ with finite support modulo Γ. It follows from (1.2.3) and simple graph-theoretical arguments that $\underline{H}_!(\mathcal{T},\mathbb{Z})^\Gamma$ is free Abelian of rank $g = g(\Gamma)$, and is a direct factor of the free Abelian group $\underline{H}(\mathcal{T},\mathbb{Z})^\Gamma$ of rank $g + c - 1$. In fact, there is a canonical injection with finite p-free cokernel (*loc. cit.* sect. 3, 6)

$$j : H_1(\Gamma \setminus \mathcal{T},\mathbb{Z}) \xrightarrow{\cong} \overline{\Gamma} \hookrightarrow \underline{H}_!(\mathcal{T},\mathbb{Z})^\Gamma,$$

which turns out to be bijective in important cases.

(1.4) *A holomorphic theta function* for Γ is an invertible holomorphic function $f : \Omega \longrightarrow C$ that for each $\gamma \in \Gamma$ satisfies

$$f(\gamma z) = c_f(\gamma)\ f(z)$$

with some $c_f(\gamma) \in C^*$, and is holomorphic non-zero at the cusps of Γ ([10] 5.1). For *meromorphic theta* functions, we allow poles and zeros on Ω, but not at the cusps. The homomorphism $c_f : \Gamma \longrightarrow \Gamma^{ab} \longrightarrow C^*$ that maps γ to $c_f(\gamma)$ is the *multiplier* of the (holomorphic or meromorphic) theta function f. The main construction of such functions is as follows. Let ω, η be fixed elements of Ω, and put

$$\theta_\Gamma(\omega,\eta,z) = \prod_{\gamma\in\tilde{\Gamma}} \left(\frac{z-\gamma\omega}{z-\gamma\eta}\right). \tag{1.4.1}$$

Note that the product is not over Γ but over its quotient $\tilde{\Gamma}$ by its center (the latter being isomorphic with a subgroup of $A^* = \mathbb{F}_q^*$), which acts effectively on Ω. The next theorem collects the principal properties of the θ_Γ. In the case of congruence subgroups Γ, it is the synopsis of several results proved in [10], mainly Thm. 5.4.1, Thm. 5.4.12, Thm. 5.7.1 and their corollaries. The reader will easily convince himself that the arguments given there apply verbatim to the case of general arithmetic subgroups as defined in (1.2).

1.5 THEOREM. (i) *The product* (1.4.1) *for* $\theta(\omega,\eta,z) = \theta_\Gamma(\omega,\eta,z)$ *converges locally uniformly* (*loc. cit.* (5.2.2)) *in* $z \in \Omega$ *and defines a meromorphic theta function for* Γ. *It is invertible (holomorphic nowhere zero) if the orbits* $\Gamma\omega$, $\Gamma\eta$ *agree, and has its only zeroes and poles at* $\Gamma\omega$, $\Gamma\eta$, *of order* $\sharp\tilde{\Gamma}_\omega$, $\sharp\tilde{\Gamma}_\eta$, *respectively, if* $\Gamma\omega \neq \Gamma\eta$.
(ii) *The multiplier* $c(\omega,\eta,\cdot) : \Gamma \longrightarrow C$ *of* $\theta(\omega,\eta,\cdot)$ *factors through* $\overline{\Gamma}$.
(iii) *Given* $\alpha \in \Gamma$, *the holomorphic theta function* $u_\alpha(z) = \theta(\omega,\alpha\omega,z)$ *is well-defined independently of* $\omega \in \Omega$, *and depends only on the class of* α *in* $\overline{\Gamma}$. *Further,* $u_{\alpha\beta} = u_\alpha u_\beta$.
(iv) $c(\omega,\eta,\alpha) = \frac{u_\alpha(\eta)}{u_\alpha(\omega)}$, *and in particular, is holomorphic in* ω *and* η.
(v) *Let* $c_\alpha(\ \cdot\ ) = c(\omega,\alpha\omega,\cdot)$ *be the multiplier of* u_α. *The rule* $(\alpha,\beta) \longmapsto c_\alpha(\beta)$ *defines*

a symmetric bilinear map on $\overline{\Gamma} \times \overline{\Gamma}$*, which takes its values in* $K_\infty^* \hookrightarrow C^*$.

(vi) *Let* $v_\infty : K_\infty^* \longrightarrow \mathbb{Z}$ *be the valuation and* $(\alpha, \beta) := -v_\infty(c_\alpha(\beta))$. *Then* $(.,.) : \overline{\Gamma} \times \overline{\Gamma} \longrightarrow \mathbb{Z}$ *is positive definite.*

As a consequence of (vi), the map $\overline{c} : \overline{\Gamma} \longrightarrow \mathrm{Hom}(\overline{\Gamma}, C^*)$ induced by $\alpha \longmapsto c_\alpha$ is injective, and the analytic group variety $\mathrm{Hom}(\overline{\Gamma}, C^*)/\overline{c}(\overline{\Gamma})$ carries the structure of an Abelian variety J_Γ defined over K_∞.

1.6 Theorem ([10] Thm. 7.4.1). J_Γ *equals the Jacobian variety of the curve* $\overline{M}_\Gamma$*, and the Abel-Jacobi map with base point* $[\omega] \in \Gamma \setminus \Omega = M_\Gamma(C)$ *is given by* $[\eta] \longmapsto$ *class of* $c(\omega, \eta, \cdot)$ *modulo* $\overline{c}(\overline{\Gamma})$.

Again, the proof given in *loc. cit.* (including its ingredients (6.5.4) and (6.4.4) carries over to the case of a general arithmetic Γ.

2. Theta functions with degenerate parameters.

(2.1) We show how functions $\theta_\Gamma(\omega, \eta, z)$ with similar properties can be defined when the parameters ω, η are allowed to take values in

$$\overline{\Omega} = \Omega \cup \mathbb{P}^1(K). \tag{2.1.1}$$

Here Γ is any arithmetic subgroup of $\mathrm{GL}(2, K)$ and $\tilde{\Gamma} \hookrightarrow \mathrm{PGL}(2, K)$ its factor group modulo the center. For $\omega, \eta \in \overline{\Omega}$ we define the rational function $F(\omega, \eta, z)$ in $z \in \mathbb{P}^1(C)$ as

$$\begin{array}{cl} \frac{z-\omega}{z-\eta} & \text{if } \omega \neq \infty \neq \eta \\ (1-\frac{z}{\eta})^{-1} & \text{if } \omega = \infty,\ \eta \neq 0, \infty \\ 1-\frac{z}{\omega} & \text{if } \eta = \infty,\ \omega \neq 0, \infty \\ z^{-1} & \text{if } \omega = \infty,\ \eta = 0 \\ z & \text{if } \eta = \infty,\ \omega = 0 \\ 1 & \text{if } \omega = \eta = \infty. \end{array} \tag{2.1.2}$$

Hence, up to cancelling, $F(z) = F(\omega, \eta, z)$ has a simple zero at ω, a simple pole at η, and is normalized such that $F(\infty) = 1$ (resp. $F(0) = 1$, resp. $F(1) = 1$) whenever the first of these conditions makes sense. We further put

$$\theta_\Gamma(\omega, \eta, z) = \prod_{\gamma \in \tilde{\Gamma}} F(\gamma\omega, \gamma\eta, z), \tag{2.1.3}$$

which specializes to (1.4.1) if both ω and η are in Ω.

(2.2) Our first task will be to establish the locally uniform convergence of the product. We let "$|\,.\,|$": $C \longrightarrow \mathbb{R}_{\geq 0}$ be the extension of the normalized absolute value on K_∞ to C and "$|\,.\,|_i$": $C \longrightarrow \mathbb{R}_{\geq 0}$ the "imaginary part" map, i.e., $|z|_i = \inf\{|z-x| \mid x \in K_\infty\}$. Besides several obvious properties, it also satisfies

$$|\gamma z|_i = \frac{\det\ \gamma}{|cz+d|^2}|z|_i \tag{2.2.1}$$

for $z \in \Omega$, $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}(2, K_\infty)$. We will perform the relevant estimates on the sets

$$U_n = \{z \in \Omega \mid |z| \leq q_\infty^n,\ |z|_i \geq q_\infty^{-n}\}. \tag{2.2.2}$$

These are affinoid subsets of $\mathbb{P}^1(C)$, and $\Omega = \bigcup_{n \in \mathbb{N}} U_n$ is an admissible covering.

2.3 Proposition. *Let $\omega, \eta \in \overline{\Omega}$ be fixed. The product (2.1.3) for $\theta_\Gamma(\omega,\eta,z)$ converges locally uniformly for $z \in \Omega$ and defines a meromorphic function on Ω. If both ω, η are in $\mathbb{P}^1(K)$ or if $\Gamma_\omega = \Gamma_\eta$, it is even invertible on Ω. Otherwise, $\theta_\Gamma(\omega,\eta,z)$ has zeroes of order $\sharp\tilde{\Gamma}_\omega$ at $\Gamma\omega$, poles of order $\sharp\tilde{\Gamma}_\eta$ at $\Gamma\eta$, and no further zeroes or poles on Ω.*

Proof. It is easily seen that the assertion is stable under replacing Γ by a commensurable group. Since any Γ is commensurable with $\mathrm{GL}(2,A)$, we may assume $\Gamma = \mathrm{GL}(2,A)$. Now for $\omega, \eta \in \Omega$, the result is [10] Prop. 5.2.3. Hence suppose that at least one of ω and η lies in $\mathbb{P}^1(K)$. Without restriction, $\omega \in \mathbb{P}^1(K)$, $\omega \neq \eta$, and $\omega \neq \infty \neq \eta$. We need the following facts, which result from (2.2.1) and/or elementary calculations:

(2.3.1) $$\{\gamma \in \Gamma \mid \gamma U_n \cap U_n \neq \emptyset\} \text{ is finite for each } n \in \mathbb{N};$$

(2.3.2) $$\frac{z-\gamma\omega}{z-\gamma\eta} - 1 = \frac{(\det\gamma)(\eta-\omega)}{(z-\gamma\eta)(c\omega+d)(c\eta+d)} \quad \left(\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma,\ \gamma\omega \neq \infty \neq \gamma\eta\right);$$

(2.3.3) $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $\gamma' = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right)$ define the same element in $\Gamma_\infty \setminus \Gamma$ if and only if $(c',d') = u(c,d)$ with some $u \in \mathbb{F}_q^*$;

(2.3.4) $$|z-\gamma\eta| \geq q_\infty^{-n} \text{ whenever } z \in U_n,\ \gamma\eta \notin U_n.$$

Combining (2.3.1) and (2.3.4) yields the existence of $c_1(n,\omega,\eta) > 0$ such that

(2.3.5) $$\frac{|\det\ \gamma|\ |\eta-\omega|}{|z-\gamma\eta|} \leq c_1(n,\omega,\eta)$$
uniformly on U_n for almost all $\gamma \in \Gamma$.

In view of (2.3.2), we must estimate $|(c\omega+d)(c\eta+d)|$ from below.

2.3.6 Claim. For given $c_2 > 0$, the number of classes of pairs (c,d) as in (2.3.3) (i.e., of classes of $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ in $\Gamma_\infty \setminus \Gamma$) such that $|(c\omega+d)(c\eta+d)| < c_2$ holds, is finite.

Proof of claim. First, exclude the finite (!) number of pairs (c,d) with $c\omega+d = 0$ or $c\eta+d = 0$. There exists $c_3(\omega) > 0$ such that the non-vanishing elements $c\omega+d$ of the fractional ideal $A\omega + A \subset K$ satisfy

(2.3.7) $$|c\omega+d| \geq c_3(\omega).$$

Hence, if $\eta \in \Omega$, the claim follows from:

(2.3.8) For any $c_4 > 0$, the number of pairs (c,d) with $|c\eta+d| < c_4$ is finite.

If $\eta \in K$, we consider the map $(c,d) \longmapsto (c\omega+d, c\eta+d)$ from $A \times A$ to $K_\infty \times K_\infty$, which by $\omega \neq \eta$ is injective. Its image is an A-lattice, which implies:

(2.3.9) Given $c_5, c_6 > 0$, the simultaneous inequalities $|c\omega+d| \leq c_5$, $|c\eta+d| \leq c_6$ are possible for a finite number of pairs only.

Since the possible values of $|c\omega+d|$, $|c\eta+d|$ are discrete and bounded from below (cf. (2.3.7)), the assertion (2.3.6) follows.

Next we observe:

$$\text{(2.3.10)}\qquad \begin{array}{l}\text{If } (c,d) \text{ as above, } n\in\mathbb{N} \text{ and } c_7>0 \text{ are fixed,}\\ \text{then } |z-\gamma\omega|\geq c_7 \text{ uniformly in } z\in U_n \text{ for almost all}\\ \gamma\in\Gamma \text{ of the form } \gamma=\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right).\end{array}$$

Now (2.3.2) together with (2.3.5), (2.3.6) and (2.3.10) yields the following:

$$\text{(2.3.11)}\qquad \begin{array}{l}\text{Given } \epsilon>0 \text{ and } n\in\mathbb{N}, \text{ almost all of the factors}\\ \text{of type } \frac{z-\gamma\omega}{z-\gamma\eta} \text{ that appear in (2.1.3) satisfy}\\ \qquad\qquad \left|\frac{z-\gamma\omega}{z-\gamma\eta}-1\right|<\epsilon\\ \text{uniformly in } z\in U_n.\end{array}$$

It remains to verify the analogous statement for the other factors in (2.1.3). They are of type

$$\text{(2.3.12)}\qquad \begin{array}{lcl}\text{(a)} & (1-\frac{z}{\gamma\eta})^{-1} & \text{if } \gamma\omega=\infty,\ \gamma\eta\neq 0,\infty\\ \text{(b)} & (1-\frac{z}{\gamma\omega}) & \text{if } \gamma\eta=\infty,\ \gamma\omega\neq 0,\infty\\ \text{(c)} & z^{-1} & \text{if } \gamma\omega=\infty,\ \gamma\eta=0\\ \text{(d)} & z & \text{if } \gamma\eta=\infty,\ \gamma\omega=0.\end{array}$$

Now cases (c) and (d) can occur only finitely many times since $\Gamma_\infty\cap\Gamma_0$ is finite. Cases (a) and (b) are similar, so we restrict to (b). Let γ_0 be such that $\gamma_0\eta=\infty$. The other such elements of Γ are the $\gamma\gamma_0$, where $\gamma\in\Gamma_\infty=\{\left(\begin{smallmatrix}a & b\\ 0 & d\end{smallmatrix}\right)\mid a,d\in\mathbb{F}_q^*,\ b\in A\}$. Thus we have to show that $\left(\begin{smallmatrix}a & b\\ 0 & d\end{smallmatrix}\right)\gamma_0\omega=\frac{a}{d}\gamma_0\omega+\frac{b}{d}$ tends with b to infinity in absolute value, which is clear. Hence the product (2.1.3) converges uniformly on each U_n to a meromorphic function with the asserted divisor. □

From now on, we omit the subscript Γ in $\theta(\omega,\eta,z)=\theta_\Gamma(\omega,\eta,z)$.

2.4 PROPOSITION. *For $\alpha\in\Gamma$, $\theta(\omega,\eta,z)$ satisfies a functional equation*

$$\theta(\omega,\eta,\alpha z)=c(\omega,\eta,\alpha)\ \theta(\omega,\eta,z)$$

with $c(\omega,\eta,\alpha)\in C^*$ independent of $z\in\Omega$.

Proof. We let $h(\omega,\eta,\alpha)$ be the quotient of $F(\omega,\eta,\alpha z)$ by $F(\alpha^{-1}\omega,\alpha^{-1}\eta,z)$. Since the two rational functions have the same divisors, $h(\omega,\eta,\alpha)$ is well-defined and constant. Now

$$\begin{aligned}\theta(\omega,\eta,\alpha z) &= \prod_{\gamma\in\tilde{\Gamma}} F(\gamma\omega,\gamma\eta,\alpha z)\\ &= \prod h(\gamma\omega,\gamma\eta,\alpha)\cdot\prod F(\alpha^{-1}\gamma\omega,\alpha^{-1}\gamma\eta,z)\\ &= \prod h(\gamma\omega,\gamma\eta,\alpha)\ \theta(\omega,\eta,z),\end{aligned}$$

whence the convergence of $c(\omega,\eta,\alpha):=\prod_{\gamma\in\tilde{\Gamma}} h(\gamma\omega,\gamma\eta,\alpha)$ results from that of $\theta(\omega,\eta,z)$, i.e., from (2.3). □

(2.5) The next step is to describe the behavior of $\theta(\omega,\eta,z)$ at the boundary, i.e., at $s \in \mathbb{P}^1(K) = \overline{\Omega} - \Omega$. As usual, possibly replacing Γ by its conjugate $\gamma\Gamma\gamma^{-1}$, where $\gamma \in \mathrm{GL}(2,K)$ satisfies $\gamma\infty = s$, it suffices to discuss the case $s = \infty$. The stabilizer $\tilde{\Gamma}_\infty$ in $\tilde{\Gamma}$ is represented by matrices $\binom{a\,b}{0\,1}$, where a runs through a subgroup W_∞ (of order w_∞, say) of $\mathbb{F}_q^*$, and b through an infinite-dimensional $\mathbb{F}_p$-vector space $\mathfrak{b} \subset K$ commensurable with a fractional A-ideal. In particular, $\mathfrak{b} \in C$ is discrete, which ensures the convergence of the infinite product written below. Put

$$t_\infty(z) = e_{\mathfrak{b}}^{-1}(z), \tag{2.5.1}$$

where $e_{\mathfrak{b}} : C \longrightarrow C$ is the function

$$e_{\mathfrak{b}}(z) = z \prod_{0 \neq b \in \mathfrak{b}} (1 - \frac{z}{b}).$$

For the essential properties of such functions, see e.g. [12] I, IV. We need the observation:

(2.5.2) $e_{\mathfrak{b}}$ is $\mathbb{F}$-linear, where $\mathbb{F} \subset \mathbb{F}_q$ is the subfield generated by W_∞. Hence for $a \in W_\infty$, $t_\infty(az) = a^{-1}t_\infty(z)$ and $t_\infty^{w_\infty}(az) = t_\infty^{w_\infty}(z)$.

It results from the fact that $\mathfrak{b}$ is even an $\mathbb{F}$-vector space.

(2.5.3) The subspace $\Omega_c = \{z \in \Omega \mid |z|_i \geq c\}$ of Ω is stable under $\tilde{\Gamma}_\infty$ and $\tilde{\Gamma}_\infty^u = \{\binom{1\,b}{0\,1} \mid b \in \mathfrak{b}\}$, and for a suitable $c \gg 0$, t_∞ identifies $\tilde{\Gamma}_\infty^u \setminus \Omega_c = \mathfrak{b} \setminus \Omega_c$ with a small pointed ball $B_\epsilon(0) - \{0\} = \{t \in C \mid 0 < |t| \leq \epsilon\}$. Again for $c \gg 0$, $\tilde{\Gamma}_\infty \setminus \Omega_c$ is an open subspace of $\Gamma \setminus \Omega \hookrightarrow \Gamma \setminus \overline{\Omega}$ (since $\gamma\Omega_c \cap \Omega_c \neq \emptyset$ implies $\gamma \in \Gamma_\infty$, cf. (2.2.1)), and $t_\infty^{w_\infty}$ is a uniformizer around the point ∞. This allows to define holomorphy, meromorphy, vanishing order at ∞, ... for functions on Ω_c invariant under $\tilde{\Gamma}_\infty^u$ or $\tilde{\Gamma}_\infty$. (For more details, see e.g. [5] V or [10] 2.7.)

As results from (2.4) and (2.3), $\theta(\omega,\eta,z)$ is invariant under $\tilde{\Gamma}_\infty^u$ and has neither zeroes nor poles on $\mathfrak{b} \setminus \Omega_c$, provided c is large (or ϵ is small) enough. It has therefore a Laurent expansion with respect to t_∞. Now the factors of type $\frac{z-\gamma\omega}{z-\gamma\eta}$ in (2.1.3) tend to 1 uniformly in γ if $|z|_i \longrightarrow \infty$, i.e., if $|t_\infty(z)| \longrightarrow 0$, hence they contribute $1+o(t_\infty)$ to the Laurent expansion. Therefore,

$$\begin{array}{l}\theta(\omega,\eta,z) \text{ is invertible around } t_\infty = 0 \text{ if} \\ \text{neither } \Gamma\omega \text{ nor } \Gamma\eta \text{ contains } \infty.\end{array} \tag{2.5.4}$$

(2.5.5) Suppose that $\infty \in \Gamma\eta \neq \Gamma\omega$. Without restriction, we may even assume $\eta = \infty$. The factors of type (b) and (d) in (2.3.12) yield

$$\prod_{\substack{\gamma \in \tilde{\Gamma}_\infty \\ \gamma\omega = 0}} z \prod_{\substack{\gamma \in \tilde{\Gamma}_\infty \\ \gamma\omega \neq 0}} (1 - \frac{z}{\gamma\omega}) = \prod_{\substack{\gamma \in \tilde{\Gamma}_\infty \\ \gamma\omega = 0}} z \prod_{\substack{\gamma \in \tilde{\Gamma}_\infty \\ \gamma\omega \neq 0}} (1 - \frac{z}{a\omega + b}),$$

writing $\gamma \in \tilde{\Gamma}_\infty$ in the form $\binom{a\,b}{0\,1}$ as above. That product defines an entire function $f : C \longrightarrow C$ with its zeroes at the points z_0 of shape $z_0 = a\omega + b$, each of the same order $\sharp\{\binom{a\,b}{0\,1} \in \tilde{\Gamma} \mid a\omega + b = z_0\}$.

Let first $\boxed{\omega \notin \mathfrak{b}}$. Since an entire function is determined up to constants by its divisor, we have, using (2.5.2):

$$\begin{aligned}
\text{const. } f(z) &= \prod_{a \in W_\infty} e_{\mathfrak{b}}(z - a\omega) \\
&= \prod_a (e_{\mathfrak{b}}(z) - a e_{\mathfrak{b}}(\omega)) \\
&= \prod_a (t_\infty^{-1}(1 + o(t_\infty))) \\
&= t_\infty^{-w_\infty}(1 + o(t_\infty)).
\end{aligned}$$

Next, let $\boxed{\omega \in \mathfrak{b}}$. Then f has zeroes of order w_∞ at the points of $\mathfrak{b}$, which gives

$$\text{const. } f(z) = e_{\mathfrak{b}}(z)^{w_\infty} = t_\infty^{-w_\infty}.$$

It is straight from definitions that for $a \in W_\infty$ (i.e., $\left(\begin{smallmatrix} a & 0 \\ 0 & 1 \end{smallmatrix}\right) \in \tilde{\Gamma}_\infty$),

$$\theta(\omega, \eta, az) = \theta(\omega, \eta, z)$$

holds. Hence, by (2.5.2), the Laurent expansion of $\theta(\omega, \eta, z)$ w.r.t. t_∞ is actually a series in $t_\infty^{w_\infty}$. Therefore, under our condition $\infty \in \Gamma\eta \neq \Gamma\omega$, $\theta(\omega, \eta, z)$ has a simple pole at the cusp represented by ∞ w.r.t. its correct uniformizer $t_\infty^{w_\infty}$. Analogous assertions hold if $\infty \in \Gamma\omega \neq \Gamma\eta$, or if $\Gamma\omega = \Gamma\eta$ (in which case the possible zeroes and poles at the cusps cancel).

We collect what has been proven.

2.6 PROPOSITION. *The function $\theta(\omega, \eta, \cdot)$ has a meromorphic continuation to the boundary $\mathbb{P}^1(K)$ of Ω. With respect to the uniformizer $t_s^{w_s}$ at the cusp $[s]$ of $\overline{M}_\Gamma$ represented by $s \in \mathbb{P}^1(K)$, it*

has a simple zero, if $s \in \Gamma\omega \neq \Gamma\eta$,
has a simple pole, if $s \in \Gamma\eta \neq \Gamma\omega$,
is invertible, if $\Gamma\omega = \Gamma\eta$ (whether or not $s \in \Gamma\omega = \Gamma\eta$). $\square$

Here of course, w_s is the *weight* of $[s]$, i.e., the size of the non-p part W_s of $\tilde{\Gamma}_s$ (cf. (2.5)).

2.7 COROLLARY. *The holomorphic function $u_\alpha(z) := \theta(\omega, \alpha\omega, z)$ on $\overline{\Omega}$ ($\omega \in \overline{\Omega}$, $\alpha \in \Gamma$ fixed) does not depend on the choice of ω.*

Proof. In view of (2.6), it suffices to verify this for $z \in \Omega$. If the parameters ω, η are in Ω, we get as in [10] Thm. 5.4.1 (iv):

$$\begin{aligned}
\frac{\theta(\omega, \alpha\omega, z)}{\theta(\eta, \alpha\eta, z)} &= \prod_{\gamma \in \tilde{\Gamma}} \left(\frac{z - \gamma\omega}{z - \gamma\alpha\omega}\right)\left(\frac{z - \gamma\alpha\eta}{z - \gamma\eta}\right) = \prod_{\gamma \in \tilde{\Gamma}} \left(\frac{z - \gamma\omega}{z - \gamma\eta}\right)\left(\frac{z - \gamma\alpha\eta}{z - \gamma\alpha\omega}\right) \\
&= \theta(\omega, \eta, z)\theta(\eta, \omega, z) = 1
\end{aligned}$$

The reader will easily verify through a case-by-case consideration that the same cancelling takes place if ω, η are allowed to take values in $\mathbb{P}^1(K)$. $\square$

2.8 DEFINITION. A *cuspidal theta function* for Γ is an invertible holomorphic function f on Ω that for each $\gamma \in \Gamma$ satisfies

$$f(\gamma z) = c_f(\gamma) f(z)$$

with some $c_f(\gamma) \in C^*$, and is meromorphic at the cusps. This means that, compared to (1.4), we allow zeroes and poles at the cusps.

The prototype of a cuspidal theta function is $\theta(\omega, \eta, \cdot)$, where both ω and η are in $\mathbb{P}^1(K)$.

2.9 Lemma. *Let $\omega, \eta \in \overline{\Omega}$, $\alpha, \gamma \in \Gamma$. The factors $F(\ .\ ,\ .\ ,\ .\ )$ of (2.1.2) satisfy*

$$\frac{F(\gamma\omega, \gamma\eta, \alpha z)}{F(\gamma\omega, \gamma\eta, z)} = \frac{F(\gamma^{-1}\alpha z, \gamma^{-1}z, \omega)}{F(\gamma^{-1}\alpha z, \gamma^{-1}z, \eta)}$$

(identity of rational functions in $z \in \mathbb{P}^1(C)$).

Proof. We may assume that $\omega \neq \eta$. Let

$$D(a,b,c,d) := \frac{a-c}{b-c} / \frac{a-d}{b-d} \quad (a,b,c,d \in \mathbb{P}^1(C))$$

be the cross-ratio which, through the usual conventions, delivers a well-defined element of $P^1(C)$ if at least three of a, b, c, d are different. Going through the cases, it is easily seen that $F(a,b,c)/F(a,b,d) = D(c,d,a,b)$, and hence the assertion follows from the invariance of $D(a,b,c,d)$ under projective transformations, in particular, under the Klein group of order 4. □

2.10 Corollary. *Let $\alpha \in \Gamma$ be fixed. The multiplier $c(\omega, \eta, \alpha)$ satisfies $c(\omega, \eta, \alpha) = \frac{u_\alpha(\eta)}{u_\alpha(\omega)}$. In particular, it is holomorphic on Ω and at the cusps, considered as a function in ω with η fixed (resp. in η with ω fixed).*

Proof. Let $\omega, \eta \in \overline{\Omega}$ be given. Then

$$\begin{aligned} c(\omega, \eta, \alpha) &= \frac{\theta(\omega, \eta, \alpha z)}{\theta(\omega, \eta, z)} = \prod_{\gamma \in \tilde{\Gamma}} \frac{F(\gamma\omega, \gamma\eta, \alpha z)}{F(\gamma\omega, \gamma\eta, z)} \\ &= \prod_{\gamma \in \tilde{\Gamma}} \frac{F(\gamma^{-1}\alpha z, \gamma^{-1}z, \omega)}{F(\gamma^{-1}\alpha z, \gamma^{-1}z, \eta)} = \frac{u_\alpha(\eta)}{u_\alpha(\omega)}, \end{aligned}$$

where the last equality follows from (2.7). □

2.11 Corollary. *Let $\omega, \eta \in \overline{\Omega}$. The constant $c(\omega, \eta, \alpha)$ and the function u_α depend only on the class of α in $\overline{\Gamma} = \Gamma^{ab}/\mathrm{tor}(\Gamma^{ab})$.*

Proof. By (2.10), the statement about $c(\omega, \eta, \alpha)$ follows from that on u_α. But $u_\alpha = \theta(\omega, \alpha\omega, \cdot)$ may be described with an arbitrary base point $\omega \in \Omega$, so the result follows from (1.5) (iii). □

2.12 Remark. As in Shimura's book [21], we may provide $\overline{\Omega}$ with a topology coming from the strong topology on $\mathbb{P}^1(C)$. To do so, it suffices to describe a fundamental system of neighborhoods for $s \in \mathbb{P}^1(K)$. By the usual homogeneity argument, we may even assume $s = \infty$, in which case the system of sets $\{\infty\} \cup \Omega_c$ $(c \in \mathbb{N})$ is as desired. It is then natural to expect that our theta functions satisfy

$$\lim_{\omega \to \omega_0,\ \eta \to \eta_0} \theta(\omega, \eta, z) = \theta(\omega_0, \eta_0, z) \tag{2.12.1}$$

with respect to that topology. This is easy to verify if e.g. all of $\omega_0, \eta_0, z \notin \Gamma\omega_0 \cup \Gamma\eta_0$ belong to Ω. On the other hand, for $\omega, \eta \in \Omega$, $\theta(\omega, \eta, z)$ is normalized such that it takes the value 1 at $z = \infty$, whereas $\theta(\infty, \eta, z)$ has a simple zero at $z = \infty$ if $\eta \notin \Gamma\infty$.

This rules out the possibility of (2.12.1) if one of the parameters ω_0, η_0 belongs to the boundary. The best we can hope for is the continuous dependence on parameters of the multiplier instead of the theta functions themselves.

2.13 Corollary. Let $\omega_0, \eta_0 \in \overline{\Omega}$, $\alpha \in \Gamma$. *Then*

$$\lim_{\omega \to \omega_0,\ \eta \to \eta_0} c(\omega, \eta, \alpha) = c(\omega_0, \eta_0, \alpha),$$

where the double limit with respect to the topology defined in (2.12) *is taken in arbitrary order.*

Proof. Apply (2.10). □

We finally note the observation, which is immediate from the product for $\theta(\omega, \eta, \cdot)$:

(2.14) The multiplier $c(\omega, \eta, \cdot) : \overline{\Gamma} \longrightarrow C^*$ has values in K_∞^* if both ω, η are in $\mathbb{P}^1(K)$.

3. Relationship with harmonic cochains.

Recall Marius van der Put's exact sequence ([24], [1])

$$0 \longrightarrow C^* \longrightarrow \mathcal{O}_\Omega(\Omega)^* \xrightarrow{r} \underline{H}(\mathcal{T}, \mathbb{Z}) \longrightarrow 0 \tag{3.1}$$

of right $\mathrm{GL}(2, K_\infty)$-modules, where the middle term is the group of invertible functions on Ω. As is explained in [10], the map r plays the role of logarithmic derivation. We briefly sketch the construction of r, and refer to *loc. cit.* for details and notations.

Let $f \in \mathcal{O}_\Omega(\Omega)^*$ and e be an oriented edge of $\mathcal{T}$ with origin v and terminus w. Then $|f|$ is constant on the rational subdomains $\lambda^{-1}(v)$ and $\lambda^{-1}(w)$ of Ω determined by v and w. Both of these are isomorphic with a projective line $\mathbb{P}^1(C)$ with $q_\infty + 1$ disjoint open balls deleted. The value of $r(f)$ on e is then

$$r(f)(e) = \log \frac{|f|_{\lambda^{-1}(w)}}{|f|_{\lambda^{-1}(v)}}, \tag{3.1.1}$$

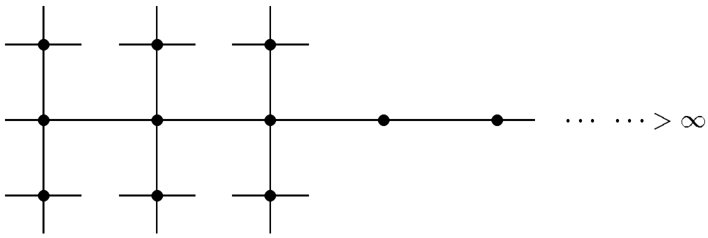
where here and in the sequel, $\log = \log_{q_\infty}$ is the logarithm to base q_∞.

Let Γ be any arithmetic subgroup of $\mathrm{GL}(2, K)$. We put $\Theta_h(\Gamma) \subset \Theta_c(\Gamma)$ for the groups of holomorphic and cuspidal theta functions for Γ as defined in (1.4) and (2.8), respectively. We have a commutative diagram

$$\begin{tikzcd}[column sep=small] & \overline{\Gamma} \arrow[dl, "\overline{u}"'] \arrow[dr, "j"] & \\ \Theta_h(\Gamma)/C^* \arrow[rr, "\overline{r}_h"] \arrow[d, hook] & & \underline{H}_!(\mathcal{T}, \mathbb{Z})^\Gamma \arrow[d, hook] \\ \Theta_c(\Gamma)/C^* \arrow[rr, "\overline{r}_c"] & & \underline{H}(\mathcal{T}, \mathbb{Z})^\Gamma, \end{tikzcd} \tag{3.2}$$

where $\overline{u}$ is derived from $\alpha \longmapsto u_\alpha$ and the horizontal maps from r. Recall that j is injective with finite prime-to-p cokernel ([10] 6.44; the proof given there applies to general arithmetic groups), and is bijective at least if $\tilde{\Gamma}$ has no prime-to-p torsion, or if K is a rational function field, ∞ the usual place at infinity, and Γ is a congruence subgroup of $\mathrm{GL}(2, A)$ [9].

(3.3) Next, we let $\mathfrak{b} \subset K$, $\tilde{\Gamma}_\infty$, $\tilde{\Gamma}_\infty^u$, $e_\mathfrak{b}$, t_∞ etc. be as in (2.5). The function $e_\mathfrak{b}$ is invertible on Ω and so $r(e_\mathfrak{b})$ is defined. The quotient graph $\tilde{\Gamma}_\infty^u \setminus \mathcal{T} = \mathfrak{b} \setminus \mathcal{T}$ has the following shape:



where the distinguished end points to ∞.

Since $r(e_\mathfrak{b}) \in \underline{H}(\mathcal{T}, \mathbb{Z})$ is invariant under $\tilde{\Gamma}_\infty^u$, it follows from the way how edges of $\mathcal{T}$ are identified mod $\mathfrak{b}$ (see e.g. proof of Proposition 3.5.1 in [10]) that for edges sufficiently close to ∞, the function $r(e_\mathfrak{b})$ grows by a factor q_∞ for each step towards ∞. In view of (3.1.1), this allows to describe the growth of $e_\mathfrak{b}(z)$ (or the decay of $t_\infty = e_\mathfrak{b}^{-1}(z)$) if $z \longrightarrow \infty$ in the topology introduced in (2.12). It is given by

$$c_1 q_\infty^{c_2|z|_i} \leq \log\ |e_\mathfrak{b}(z)| \leq c_1' q_\infty^{c_2|z|_i} \quad (|z|_i \gg 0) \tag{3.3.1}$$

for suitable constants $0 < c_1 < c_1'$, $c_2 > 0$ depending on $\mathfrak{b}$. (These constants can be made explicit if the need arises, see e.g. [7] for the case of $A = \mathbb{F}_q[T]$.) Note that multiplying z by the inverse π_∞^{-1} of a uniformizer π_∞ of K_∞ corresponds to shifting $\lambda(z)$ by one towards ∞, using again the terminology of [10].

Similarly, if $f \in \mathcal{O}_\Omega(\Omega)^*$ is invariant under $\tilde{\Gamma}_\infty^u$, its logarithmic derivative $r(f)$ may be considered as a function on edges of $\mathfrak{b} \setminus \mathcal{T}$, which implies that f must satisfy similar estimates

$$c_3 q_\infty^{c_4|z|_i} \leq \log\ |f(z)| \leq c_3' q_\infty^{c_4|z|_i}$$

for $|z|_i$ large. Hence, multiplying $f(z)$ by a suitable power t_∞^k of t_∞, the resulting $t_\infty^k f(z)$ will be bounded around $t_\infty = 0$, and $f(z)$ is meromorphic at ∞. The same reasoning applies to the other cusps. Thus:

(3.3.2) If $f \in \mathcal{O}_\Omega(\Omega)^*$ is invariant under the unipotent radical $\tilde{\Gamma}_s^u$ of $\tilde{\Gamma}_s$ then f is meromorphic at the cusp represented by $s \in \mathbb{P}^1(K)$.

3.4 PROPOSITION. *The maps $\overline{r}_h$ and $\overline{r}_c$ in* (3.2) *are bijective.*

Proof. For $\overline{r}_h$, this is [10] 6.4.3. Injectivity of $\overline{r}_c$ follows directly from (3.1). Thus let $\varphi \in \underline{H}(\mathcal{T}, \mathbb{Z})^\Gamma$ equal $r(f)$ with $f \in \mathcal{O}_\Omega(\Omega)^*$. Then f satisfies $f(\gamma z) = c_f(\gamma) f(z)$ for $\gamma \in \Gamma$. The map $\gamma \longmapsto c_f(\gamma)$ is a homomorphism, which vanishes on p-groups of type $\tilde{\Gamma}_s^u$. By (3.3.2), f is meromorphic at the cusps, and is therefore a cuspidal theta function. $\square$

(3.5) We let $\Theta_c'(\Gamma) \subset \Theta_c(\Gamma)$ be the subgroup of cuspidal theta functions f whose multiplier $c_f : \tilde{\Gamma}^{ab} \longrightarrow C^*$ factors over $\overline{\Gamma} = \Gamma^{ab}/\mathrm{tor}(\Gamma^{ab}) = \tilde{\Gamma}^{ab}/\mathrm{tor}(\tilde{\Gamma}^{ab})$. Since the prime-to-$p$ torsion of $\tilde{\Gamma}^{ab}$ is always finite ([20] II, sect. 2, Ex. 2), the inclusion

$$\begin{array}{rcl} \Theta_c(\Gamma)/\Theta_c'(\Gamma) & \hookrightarrow & \mathrm{Hom}(\mathrm{tor}(\tilde{\Gamma}^{ab}), C^*) \\ f & \longmapsto & c_f \mid \mathrm{tor}(\tilde{\Gamma}^{ab}) \end{array} \tag{3.5.1}$$

shows that the index $[\Theta_c(\Gamma) : \Theta'_c(\Gamma)]$ is always finite and not divisible by p. Note that $\mathrm{Hom}(\mathrm{tor}(\tilde{\Gamma}^{ab}), C^*)$ is trivial if Γ has no prime-to-p torsion, as follows e.g. from (1.2.7) (ii). Hence $\Theta_c(\Gamma) = \Theta'_c(\Gamma)$ in this case.

3.6 Lemma. Let $j : \overline{\Gamma} \hookrightarrow \underline{H}_!(\mathcal{T}, \mathbb{Z})^\Gamma$ be the canonical inclusion. We have

$$j(\overline{\Gamma}) = \underline{H}_!(\mathcal{T}, \mathbb{Z})^\Gamma \cap r(\Theta'_c(\Gamma)).$$

Proof. The inclusion of $j(\overline{\Gamma})$ in $r(\Theta'_c(\Gamma))$ comes from (1.5) (ii), i.e., the fact that c_α factors through $\overline{\Gamma}$. The opposite inclusion is [10] Cor. 7.5.3. $\square$

(3.7) We next interpret the quotient $r(\Theta'_c(\Gamma))/j(\overline{\Gamma})$ as the group of cuspidal divisors of degree zero on the curve $\overline{M}_\Gamma$. Recall that $cusp(\Gamma) = \Gamma \setminus \mathbb{P}^1(K)$ is the set of cusps, of order $c = c(\Gamma)$, and for each $[s] \in cusp(\Gamma)$, $w_s = [\tilde{\Gamma}_s : \tilde{\Gamma}_s^u]$ is its weight. We put

$$D_\infty := D_\infty(\Gamma) := \mathbb{Z}[cusp(\Gamma)]$$

for the group of cuspidal divisors on $\overline{M}_\Gamma$. At $[s]$, each $f \in \Theta_c(\Gamma)$ has an expansion w.r.t. t_s, and even w.r.t. $t_s^{w_s}$ if $f \in \Theta'_c(\Gamma)$. We let $\mathrm{ord}_{[s]}(f)$ be the order of f w.r.t. t_s (which clearly depends only on the class $[s]$ of s) and

(3.7.1) $$\mathrm{div}(f) = \sum_{[s] \in cusp(\Gamma)} \frac{\mathrm{ord}_{[s]} f}{w_s} [s] \in D_\infty \otimes \mathbb{Q}.$$

3.8 Theorem. *The map $f \longmapsto \mathrm{div}(f)$ induces an isomorphism*

$$\overline{\mathrm{div}} : r(\Theta'_c(\Gamma))/j(\overline{\Gamma}) \xrightarrow{\cong} D^0_\infty,$$

where $D^0_\infty \hookrightarrow D_\infty$ is the subgroup of divisors of zero degree.

Proof. For $f \in \Theta'_c(\Gamma)$, $\mathrm{div}(f)$ lies in D_∞, as follows from (2.5.2). Clearly, div restricted to $\underline{H}_!(\mathcal{T}, \mathbb{Z})^\Gamma$ (or more precisely, to those f such that $r(f) \in \underline{H}_!(\mathcal{T}, \mathbb{Z})^\Gamma$) is trivial, hence $\overline{\mathrm{div}}$ is well-defined. It is surjective by (2.6) and injective since, by (3.4) and (3.6), $r(\Theta'_c(\Gamma)/j(\overline{\Gamma})$ is free Abelian of rank $c(\Gamma) - 1$. $\square$

3.9 Corollary. *$\Theta'_c(\Gamma)$ is the group generated by the constants C^* and the functions $\theta(\omega, \eta, \cdot)$ with $\omega, \eta \in \mathbb{P}^1(K)$.*

Proof. Obvious from (3.8), (3.6), (3.4), and (2.11). $\square$

For what follows, we write Θ'_c for $\Theta'_c(\Gamma)$, and abbreviate $\underline{H}(\mathcal{T}, \mathbb{Z})^\Gamma$ and $\underline{H}_!(\mathcal{T}, \mathbb{Z})^\Gamma$ by $\underline{H}$ and $\underline{H}_!$, respectively. Let l be the least common multiple of the weights w_s, $[s] \in cusp(\Gamma)$.

3.10 Corollary. The index of $(\underline{H}_! + r(\Theta'_c))/\underline{H}_! \xrightarrow{\cong} r(\Theta'_c)/j(\overline{\Gamma}) \xrightarrow{\cong} D^0_\infty$ in $\underline{H}/\underline{H}_!$ is a divisor of $l^{-1} \prod_{[s] \in cusp(\Gamma)} w_s$, and the quotient group is annihilated by $q - 1$.

Proof. We may extend $\overline{\mathrm{div}}$ to a map from $\underline{H}/\underline{H}_!$ into the elements of degree zero of $\oplus_{[s]} w_s^{-1}\mathbb{Z}[s] \hookrightarrow D_\infty \otimes \mathbb{Q}$. The inverse image of D^0_∞ is precisely $(\underline{H}_! + r(\Theta'_c))/\underline{H}_!$, as follows from (3.8). The assertion now results from chasing in the diagram

(3.10.1) $$\begin{array}{ccccccccc} 0 & \longrightarrow & D^0_\infty & \longrightarrow & D_\infty & \xrightarrow{\deg} & \mathbb{Z} & \longrightarrow & 0 \\ & & \cap & & \cap & & \cap & & \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & (\oplus w_s^{-1}\mathbb{Z}[s])^0 & \longrightarrow & \oplus w_s^{-1}\mathbb{Z}[s] & \longrightarrow & l^{-1}\mathbb{Z} & \longrightarrow & 0 \end{array}$$

and noting that the w_s are divisors of $q-1$. $\square$

(3.11) Since $\underline{H}_!$ is a space of functions with finite support on the edges of the graph $\Gamma \setminus \mathcal{T}$, it is provided with a natural bilinear form

$$(.,.): \underline{H}_! \times \underline{H}_! \longrightarrow \mathbb{Q}.$$

If $\tilde{\Gamma}_e$ is the stabilizer of $e \in Y(\mathcal{T})$, the volume of the corresponding edge of $\Gamma \setminus \mathcal{T}$ is $\frac{1}{2}\sharp(\tilde{\Gamma}_e)^{-1}$. Two remarks are in order.

(3.11.1) $(.,.)$ as defined above is the restriction of the Petersson scalar product on $\underline{H}_!(\mathcal{T},\mathbb{C})^{\Gamma}$, which is a space of automorphic forms. In fact, the restriction of $(.,.)$ to $\overline{\Gamma} \xrightarrow{\cong} j(\overline{\Gamma}) \hookrightarrow \underline{H}_!$ agrees with the pairing $(.,.)$ in (1.5) (vi) ([10] 5.7.1), and in particular, takes its values in $\mathbb{Z}$.

(3.11.2) There exists a natural extension of $(.,.)$ to a pairing labeled by the same symbol

$$(.,.): \underline{H}_! \times \underline{H} \longrightarrow \mathbb{Q}.$$

It is characterized through its restriction to $j(\overline{\Gamma}) \times r(\Theta'_c)$, where it satisfies

$$(r(u_\alpha), r(f)) = -v_\infty(c_f(\alpha)), \tag{3.11.3}$$

compare (3.2) and (1.5) (vi). Finally, we put

$$\underline{H}_!^{\perp} := \{\varphi \in \underline{H}(\mathcal{T},\mathbb{Z})^{\Gamma} \mid (\underline{H}_!, \varphi) = 0\}. \tag{3.11.4}$$

Then $\underline{H}_!^{\perp}$ is a direct factor of $\underline{H}$ and "almost complementary" to $\underline{H}_!$, i.e., $\underline{H}/\underline{H}_! \oplus \underline{H}_!^{\perp}$ is finite. We will see at once that this group is closely related to the cuspidal divisor class group of $\overline{M}_\Gamma$.

4. The cuspidal divisor class group.

From now on, we assume that Γ is a congruence subgroup of some $\mathrm{GL}(Y)$. The next result follows from determining the divisors of certain modular units (analogues of classical Weber or Fricke functions) and expressing them through partial zeta functions. This has been carried out in detail in the special cases where

A) the base ring A is a polynomial ring $\mathbb{F}_q[T]$ and $\Gamma \subset \mathrm{GL}(2,A)$ is an arbitrary congruence subgroup [2], or

B) the base ring A is subject only to the conditions given in (1.1), but $\Gamma = \mathrm{GL}(Y)$ is the full linear group of a rank-two A-lattice Y [5].

The proof of the general case (A and Γ without further restrictions) will follow e.g. by combining the methods of [2] and [5]. The necessary ingredients are sketched in [5] VI.5.13, but still some work has to be done to complete the argument. A rather short proof which avoids the difficult calculations of *loc. cit.* will be given in [8].

4.1 Theorem. *Let Γ be a congruence subgroup of $GL(2,K)$. The cuspidal divisors of degree zero on $\overline{M}_\Gamma$ generate a finite subgroup $\mathcal{C}(\Gamma)$ of the Jacobian J_Γ of $\overline{M}_\Gamma$.*

The corresponding result for classical modular curves has been proven by Manin and Drinfeld [14]; a different proof has been given by Kubert and Lang [13]. Our aim is now to give a more accurate description of $\mathcal{C} = \mathcal{C}(\Gamma)$.

4.2 PROPOSITION. *Let f be a modular unit, i.e., a meromorphic function on $\overline{M}_\Gamma$ with its divisor supported by the cusps. Then $r(f) \in \underline{H}_!^\perp = \underline{H}_!(\mathcal{T},\mathbb{Z})^\Gamma)^\perp$. Conversely, if $f \in \Theta_c'(\Gamma)$ is such that $r(f) \in \underline{H}_!^\perp$ then $f^{q_\infty - 1}$ is a modular unit.*

Proof. Since f is invertible on Ω, $r(f)$ is defined, and $r(f) \in \underline{H}_!^\perp$ follows from (3.11.3). Let $f \in \Theta_c'$ be such that $r(f) \in \underline{H}_!^\perp$, and let $\chi = c_f$ be its multiplier. By (4.1) there exists $n \in \mathbb{N}$ and a modular unit g such that f^n/g is holomorphic on Ω and at the cusps. From [10] 7.5.3 $f^n/g = \text{const.}\ u_\alpha$ for some $\alpha \in \Gamma$, hence $\chi^n = c_\alpha$. Since $r(f) \perp j(\overline{\Gamma})$, we have $|c_\alpha(\beta)| = 1$ for all $\beta \in \Gamma$, which gives $c_\alpha = 1$. Therefore, χ has finite order, which by (3.9) and (2.14) is a divisor of $q_\infty - 1$. $\square$

(4.3) We let P_∞ be the divisors of modular units, i.e., the principal divisors on $\overline{M}_\Gamma$ supported by the cusps. The map $\operatorname{div}(f) \longmapsto r(f)$ identifies P_∞ with a subgroup of $\underline{H}_!^\perp \hookrightarrow \underline{H}$, which by abuse of language will be labeled by the same symbol P_∞. By the above,

$$(q_\infty - 1)(\underline{H}_!^\perp \cap r(\Theta_c')) \subset P_\infty \subset \underline{H}_!^\perp \cap r(\Theta_c'), \tag{4.3.1}$$

and the group $\mathcal{C}$ of cuspidal divisor classes is

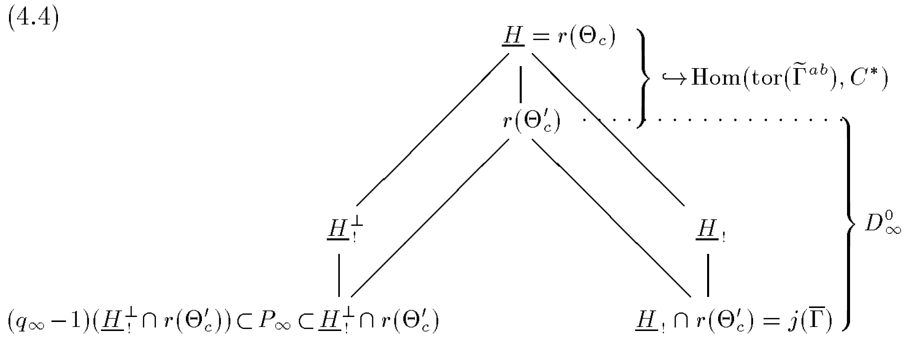
$$\mathcal{C} = D_\infty^0/P_\infty \xrightarrow{\cong} r(\Theta_c')/(j(\overline{\Gamma}) \oplus P_\infty). \tag{4.3.2}$$

We therefore have an exact sequence

$$0 \longrightarrow U \longrightarrow \mathcal{C} \longrightarrow V \longrightarrow 0, \tag{4.3.3}$$

where $U = \underline{H}_!^\perp \cap r(\Theta_c')/P_\infty$ is isomorphic with a quotient of $(\mathbb{Z}/(q_\infty - 1)\mathbb{Z})^{c(\Gamma)-1}$ and $V = r(\Theta_c')/(j(\overline{\Gamma}) \oplus \underline{H}_!^\perp \cap r(\Theta_c')) \hookrightarrow \underline{H}/\underline{H} \oplus \underline{H}_!^\perp$. The following diagram displays the inclusions.

(4.4)



4.5 REMARKS. (i) As follows from (1.2.7), the vertical inclusions are bijective if Γ has no non-p torsion, in which case $V = \underline{H}/\underline{H}_! \oplus \underline{H}_!^\perp$.
(ii) In the general case, both U and the cokernel of V in $\underline{H}/\underline{H}_! \oplus \underline{H}_!^\perp$ have prime-to-p order. Hence the p-parts of $\mathcal{C}$ and of $\underline{H}/\underline{H}_! \oplus \underline{H}_!^\perp$ always agree.
(iii) We know of no single example of a congruence group Γ such that $j(\overline{\Gamma}) \neq \underline{H}_! = \underline{H}_!(\mathcal{T},\mathbb{Z})^\Gamma$. The two groups agree at least if $A = \mathbb{F}_q[T]$ (see [9]). However, there are examples, given in the next section, where $r(\Theta_c')$ and even $r(\Theta_c') + \underline{H}_!$ differs from $\underline{H}$.

The description for the Jacobian J_Γ of $\overline{M}_\Gamma$ given in (1.6) is valid over each complete subextension of C/K_∞, in particular, over K_∞ itself. We let $\phi_\infty(\Gamma)$ be the group of connected components of the Néron model $\mathcal{J}_\Gamma$ of J_Γ/K_∞.

4.6 Theorem. *$\phi_\infty(\Gamma)$ is canonically isomorphic with* $\mathrm{Hom}(\overline{\Gamma}, \mathbb{Z})/i(\overline{\Gamma})$, *where* $i : \overline{\Gamma} \hookrightarrow \mathrm{Hom}(\overline{\Gamma}, \mathbb{Z})$ *comes from the pairing* $(.,.)$ *on* $\overline{\Gamma}$.

Proof. Easy consequence of the construction of J_Γ ([10] sect. 7) and Mumford's results [17] on degenerating Abelian varieties. Details are given in [6] Cor. 2.11. The assumption of $A = \mathbb{F}_q[T]$ made in that paper is not used in an essential fashion. □

There is a canonical map can_∞ from $\mathcal{C} = \mathcal{C}(\Gamma)$ to $\phi_\infty(\Gamma)$, which to each divisor class $[D]$ associates the component of the reduction of $[D]$ at infinity. Combining what we know about these groups ((4.3), (4.4), (4.6)) yields the following description of can_∞.

4.7 Corollary. The map $\mathrm{can}_\infty : \mathcal{C}(\Gamma) \longrightarrow \phi_\infty(\Gamma)$ is given by

$$\begin{array}{rcl} \mathcal{C}(\Gamma) \xrightarrow{\cong} r(\Theta_c')/(j(\overline{\Gamma}) \oplus P_\infty) & \longrightarrow & \mathrm{Hom}(\overline{\Gamma}, \mathbb{Z})/i(\overline{\Gamma}) \xrightarrow{\cong} \phi_\infty(\Gamma) \\ \text{class of } r(f) & \longmapsto & \text{class of } (-v_\infty \circ c_f). \end{array}$$

Here $c_f : \overline{\Gamma} \longrightarrow K_\infty^*$ is the multiplier of f and $v_\infty : K_\infty^* \longrightarrow \mathbb{Z}$ the valuation.

Obviously, the kernel of can_∞ is $j(\overline{\Gamma}) \oplus (\underline{H}_!^\perp \cap r(\Theta_c'))/j(\overline{\Gamma}) \oplus P_\infty$, i.e., the group U of (4.3.3). As we will see, can_∞ need neither be injective nor surjective.

We finally recall the fact that each congruence subgroup Γ' contains a congruence subgroup Γ without prime-to-p torsion. For such Γ, (4.5) (i) applies, and (4.7) becomes

$$\mathcal{C}(\Gamma) \xrightarrow{\cong} \underline{H}/\underline{H}_! \oplus P_\infty \xrightarrow{\text{proj.}} \underline{H}/\underline{H}_! \oplus \underline{H}_!^\perp \hookrightarrow \mathrm{Hom}(\underline{H}_!, \mathbb{Z})/i(\underline{H}_!) \xrightarrow{\cong} \phi_\infty(\Gamma). \tag{4.8}$$

Hence in this case, $\phi_\infty(\Gamma)$ as well as the image $\phi_\infty^{cusp}(\Gamma) := \mathrm{can}_\infty(\mathcal{C}(\Gamma))$ of the cuspidal divisor classes may be described entirely in terms of the almost finite graph $\Gamma \setminus \mathcal{T}$. Note that assertions similar to (4.6) - (4.8) are valid also in the case of a general arithmetic group Γ (i.e., without the assumption of being a congruence subgroup), except for the finiteness of $\mathcal{C}(\Gamma)$. By analogy with the number field case [18], that latter is unlikely to hold.

5. The case of Hecke congruence subgroups over a polynomial ring.

We now assume that A equals the polynomial ring $\mathbb{F}_q[T]$ and Γ is the Hecke congruence subgroup $\Gamma_0(n) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}(2, A) \mid c \equiv 0 \mod n\}$ for a certain $n \in A$. A lot of material about these groups, including structural properties of $\Gamma \setminus \mathcal{T}$, formulae for $g(\Gamma)$, $c(\Gamma)$ etc., may be found in [9]. Note in particular that (*loc. cit.*, Thm. 3.3)

$$H_1(\Gamma \setminus \mathcal{T}, \mathbb{Z}) \cong \overline{\Gamma} \xrightarrow[j]{\cong} \underline{H}_! = \underline{H}_!(\mathcal{T}, \mathbb{Z})^\Gamma. \tag{5.1}$$

(5.2) We start with a few examples that illustrate how $\mathrm{can}_\infty : \mathcal{C}(\Gamma) \longrightarrow \phi_\infty(\Gamma)$ may be calculated. Let $q = 2$. Apart from the general advantage that $g(\Gamma)$ and $c(\Gamma)$ are then small, $q = 2$ forces that

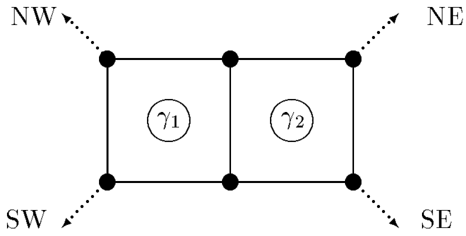
(5.2.1) the group U of (4.3.3) is trivial, hence

(5.2.2) $\mathrm{can}_\infty : \mathcal{C}(\Gamma)$ is injective, and

(5.2.3) $\mathcal{C}(\Gamma) = \underline{H}/\underline{H}_! \oplus \underline{H}_!^\perp$, due to (3.10).

5.3 Examples.

(5.3.1) $\Gamma = \Gamma_0(n)$, $n = T(T^2 + T + 1) \in \mathbb{F}_2[T]$. The graph $\Gamma \setminus \mathcal{T}$ looks:



Here $\cdots >$ indicates a cusp. Let γ_1, γ_2 be the two cycles of length 4, oriented counter-clockwise, and $\varphi_1, \varphi_2, \varphi_3$ the $\mathbb{Z}$-valued harmonic cochains flowing from the SW, the SE, the NE cusp, respectively, to the NW cusp, going the way round counter-clockwise. Then $\{\gamma_1, \gamma_2\}$ and $\{\gamma_1, \gamma_2, \varphi_1, \varphi_2, \varphi_3\}$ are $\mathbb{Z}$-bases of $\underline{H}_!$ and $\underline{H}$, respectively. With respect to these bases, the pairing $(.,.) : \underline{H}_! \times \underline{H} \longrightarrow \mathbb{Z}$ is given by

	γ_1	γ_2	φ_1	φ_2	φ_3
γ_1	4	-1	2	1	1
γ_2	-1	4	3	2	1

.

We get $\sharp\phi_\infty(\Gamma) = \left|\det \begin{pmatrix} 4, & -1 \\ -1, & 4 \end{pmatrix}\right| = 15$, and after an elementary computation, $\sharp\mathcal{C}(\Gamma) = [\underline{H} : \underline{H}_! \oplus \underline{H}_!^{\perp}] = 15$, too. Hence can_∞ is bijective.

N.B. J_Γ splits into two elliptic curves with 3 resp. 5 rational points over $K = \mathbb{F}_2(T)$, which are therefore all "cuspidal" ([6] 4.4).

(5.3.2) Drawings of the graphs $\Gamma \setminus \mathcal{T}$ ($\Gamma = \Gamma_0(n)$) for the next examples may be found in [19]. For these, the matrix of $(.,.) : \underline{H}_! \times \underline{H} \longrightarrow \mathbb{Z}$ and thus $\mathcal{C}$ and ϕ_∞ may be calculated as above. We restrict to giving the results. In all cases, can_∞ is bijective (which, however, is not typical: see (5.3.3)!).

$n \in \mathbb{F}_2[T]$	$g(\Gamma)$	$c(\Gamma)$	$\mathcal{C}(\Gamma)$
$T^2(T+1)$	1	6	$\mathbb{Z}/6\mathbb{Z}$
T^3	1	4	$\mathbb{Z}/4\mathbb{Z}$
T^3+T+1	2	2	$\mathbb{Z}/7\mathbb{Z}$
$(T^2+T+1)^2$	2	5	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/10\mathbb{Z}$
T^4	3	6	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$

(5.3.3) $\Gamma = \Gamma_0(n)$, where (i) $n = T^4 + T^3 + 1$ or (ii) $n = T^4 + T + 1$, which both are irreducible over $\mathbb{F}_2$. In both cases, $g(\Gamma) = 4$, $c(\Gamma) = 2$, $\sharp\mathcal{C}(\Gamma) = 5$ (see also (5.6)). However, $\phi_\infty(\Gamma) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/80\mathbb{Z}$ for (i) and $\phi_\infty(\Gamma) \cong \mathbb{Z}/45\mathbb{Z}$ for (ii). Hence can_∞ is *not* surjective in these cases.

We let now again $\mathbb{F}_q$ be an arbitrary finite field, n a monic polynomial of degree d in $A = \mathbb{F}_q[T]$, and $\Gamma = \Gamma_0(n)$. We give an intrinsic description of the group $\Theta_c'(\Gamma)$ of (3.5).

5.4 Theorem. *Let n have h different monic prime divisors in A. Then $\Theta_c'(\Gamma)$ has index $(q-1)^{2^{h-1}}$ in $\Theta_c(\Gamma)$.*

Proof. Without restriction, we may assume $q > 2$.

(i) By (5.1) and (3.6), $\underline{H}_! \subset r(\Theta_c')$, hence $\Theta_c/\Theta_c' \xrightarrow{\cong} \underline{H}/r(\Theta_c')$. Consider the commutative diagram with exact rows:

(5.4.1)
$$\begin{tikzcd} 0 \arrow[r] & r(\Theta'_c)/\underline{H}_! \arrow[r] \arrow[d, "\wr"] & \underline{H}/\underline{H}_! \arrow[r] \arrow[d, hook, "\overline{\mathrm{div}}"] & \underline{H}/r(\Theta'_c) \arrow[r] \arrow[d, "\alpha"] & 0 \\ 0 \arrow[r] & \mathrm{Div}^0_\infty \arrow[r] & \Big(\bigoplus\limits_{[s]\in cusp(\Gamma)} w_s^{-1}\mathbb{Z}[s]\Big)^0 \arrow[r] & \Big(\bigoplus w_s^{-1}\mathbb{Z}[s]\,/\mathbb{Z}[s]\Big)^0 \arrow[r] & 0 \end{tikzcd}$$

The right hand arrow α is injective; it suffices therefore to calculate its image.

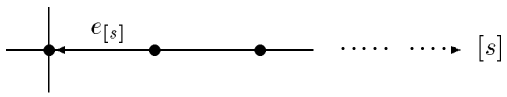
(ii) From [9] 2.15 we know that Γ has precisely 2^h cusps $[s]$ with $w_s = q-1$ (the *regular* cusps), and for the other (*irregular*) cusps, $w_s = 1$. Hence the lower right group in (5.4.1) equals $(\mathbb{Z}/(q-1)\mathbb{Z})^{reg(\Gamma),0}$, the subgroup of elements of degree zero in $(\mathbb{Z}/(q-1)\mathbb{Z})^{reg(\Gamma)}$, where $reg(\Gamma)$ is the set of regular cusps. Using this identification,

$$\alpha :\ r(\Theta_c)/r(\Theta'_c) = \underline{H}/r(\Theta'_c) \hookrightarrow (\mathbb{Z}/(q-1)\mathbb{Z})^{reg(\Gamma),0}$$

associates with each $r(f) \in \underline{H}$ the 2^h-tuple $(\ldots, \mathrm{ord}_{[s]}f \mod q-1, \ldots)$.

(iii) We have to introduce some more notation. Suppose from now on that $d :=$ deg $n \geq 2$. (The case $d=1$, which leads to $g(\Gamma)=0$, $c(\Gamma)=2$, $\Gamma \setminus \mathcal{T}$ isomorphic with a straight line $\cdots - - - \bullet - - - \bullet - - - \bullet \cdots$, is easily dealt with directly. The result follows in this case also from (5.7).)

Then to each cusp $[s]$ there corresponds a maximal half-line $hl[s]$ of $\Gamma \setminus \mathcal{T}$. We let $e_{[s]}$ be the first edge of $hl[s]$, oriented away from $[s]$, and call it the base edge of $[s]$.



5.4.3 CLAIM. For each $f \in \Theta_c$, we have $\mathrm{ord}_{[s]}f = r(f)(e_{[s]})$.

For the proof of this fact, it suffices to verify $r(t_s)(e_{[s]}) = 1$, where t_s is the corresponding uniformizer, cf. (2.5). As usual, possibly replacing Γ by a conjugate, we may assume $s = \infty$, in which case the assertion is a consequence of

- Proposition 1.14 of [7],
- the way how vertices and edges of $\mathcal{T}$ are identified under Γ_∞,

and the trivial but crucial fact:

- each fractional ideal $\mathfrak{b}$ of K has a direct complement of the form (π_∞^r) in K_∞. Here π_∞ is a uniformizer at ∞, e.g. $\pi_\infty = T^{-1}$.

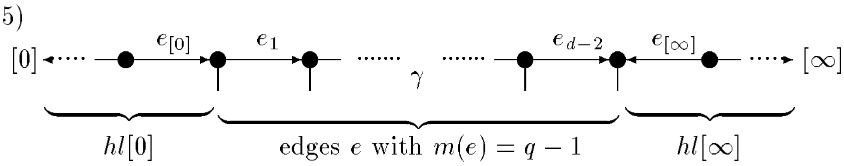
(iv) Let $\varphi \in \underline{H} = \underline{H}(\mathcal{T}, \mathbb{Z})^\Gamma$. The harmonicity condition (1.3.1) for φ as a function on $\Gamma \setminus \mathcal{T}$ reads

(5.4.4)
$$\sum_{\substack{e \in Y(\Gamma\setminus\mathcal{T}) \\ o(e)=v}} m(e)\varphi(e) = 0$$

for each vertex v of $\Gamma \setminus \mathcal{T}$, where the multiplicity $m(e)$ $(1 \leq m(e) \leq q+1)$ takes care of the identification of edges of $\mathcal{T}$ modulo Γ. Clearly, $\sum\limits_{o(e)=v} m(e) = q+1$.

(v) The next statements result from the description of $\Gamma \setminus \mathcal{T}$ given in [9]. As usual, $[0]$ and $[\infty]$ denote the cusps represented by $(0:1)$ and $(1:0)$, respectively. Their corresponding half-lines $hl[0]$ and $hl[\infty]$ in $\Gamma \setminus \mathcal{T}$ are connected by a path γ consisting of a sequence of $d-2$ edges $e_1, \ldots, e_{d-2}$ of valence 3. The edges $e = \overline{e}_{[0]}, e_1, \overline{e}_1, \ldots, e_{d-2}, \overline{e}_{d-2}, \overline{e}_{[\infty]}$ enter with multiplicity $m(e) = 1$ into (5.4.4), whereas the $d-1$ edges connecting $hl[0] \cup \gamma \cup hl[\infty]$ with the rest of $\Gamma \setminus \mathcal{T}$ have multiplicity $q-1$, always with respect to vertices on γ. This is the picture:

(5.4.5)



(vi) By the above, for any $\varphi \in \underline{H}$ we have

$$\varphi(e_{[0]}) \equiv \varphi(e_1) \equiv \cdots \equiv \varphi(e_{d-2}) \equiv -\varphi(e_{[\infty]}) \mod q-1.$$

The group W of Atkin-Lehner-involutions (which acts on $\overline{M}_\Gamma$ as well as on $\Gamma \setminus \mathcal{T}$) acts transitively on $reg(\Gamma)$, and some pair $([s],[s'])$ of regular cusps lies in the W-orbit of $([0],[\infty])$ if and only if $[s'] = w[s]$, where $w = w_n$ is the total involution induced from the matrix $\binom{0\,1}{n\,0} \in \mathrm{GL}(2,K)$. Hence for any $\varphi \in \underline{H} \in reg(\Gamma)$,

$$\varphi(e_{[s]}) \equiv -\varphi(e_{w[s]}) \mod q-1 \tag{5.4.6}$$

holds. On the other hand, it is obvious from (5.4.5) that for each pair $([s], w[s])$ of w-conjugate regular cusps there exists a harmonic cochain $\varphi \in \underline{H}$ such that $\varphi(e_{[s]}) = 1$, $\varphi(e_{w[s]}) = -1$. Hence the image of α in $(\mathbb{Z}/(q-1)\mathbb{Z})^{reg(\Gamma),0}$ (see (5.4.2)) agrees with the free $\mathbb{Z}/(q-1)\mathbb{Z}$-submodule of rank $\frac{1}{2}\sharp\, reg(\Gamma) = 2^{h-1}$ defined by the congruence condition (5.4.6), which finally yields the result. $\square$

5.5 Corollary. *With notations as in* (5.4)*, the cokernel* $\phi_\infty/\phi_\infty^{cusp}$ *of* $\mathrm{can}_\infty : \mathcal{C}(\Gamma) \longrightarrow \phi_\infty(\Gamma)$ *has order a multiple of* $(q-1)^{2^{h-1}} : [\underline{H}_!^\perp : \underline{H}_!^\perp \cap r(\Theta_c')]$.

Proof. With identifications as in (4.7), $\phi_\infty^{cusp} = r(\Theta_c')/\underline{H}_! \oplus \underline{H}_!^\perp \cap r(\Theta_c') \hookrightarrow \underline{H}/\underline{H}_! \oplus \underline{H}_!^\perp \hookrightarrow \phi_\infty$. The stated value is the index of ϕ_∞^{cusp} in $\underline{H}/\underline{H}_! \oplus \underline{H}_!^\perp$. $\square$

For the remainder of this section, we suppose in addition that n is *prime*. The cuspidal divisor class group $\mathcal{C} = \mathcal{C}(\Gamma)$ of $\Gamma = \Gamma_0(n)$ has been determined in [3] and, with different methods, in [7]. The result is

5.6 Theorem. *In the above situation,* $\mathcal{C}$ *is cyclic of order* $\frac{q^d-1}{q^2-1}$ *if* $d = \deg\, n$ *is even and* $\frac{q^d-1}{q-1}$ *if* d *is odd.*

Here $c(\Gamma) = 2$ with the two cusps $[0]$ and $[\infty]$. A meromorphic function f on $\overline{M}_\Gamma$ with divisor $\sharp(\mathcal{C})([0]-[\infty])$ may be constructed as follows. Let $\Delta : \Omega \longrightarrow C$ be the Drinfeld discriminant (see e.g. [7]) and $\Delta_n(z) = \Delta(nz)$. Then Δ/Δ_n is a modular function (i.e., invariant) for Γ and $\mathrm{div}(\Delta/\Delta_n) = (q^d-1)([0]-[\infty])$ (*loc. cit.* (3.11)). Let now

$$\begin{aligned} r \; &:= \; (q^2-1)(q-1) \quad \text{for even } d \\ &= \; (q-1)^2 \quad \text{for odd } d. \end{aligned}$$

Using the machinery of Drinfeld modular forms, it is further shown in [7] 3.18:

5.7 Theorem. *Δ/Δ_n admits an r-th root in $\mathcal{O}_\Omega(\Omega)^*$, and r is maximal with this property.*

(5.8) Let D_n be such an r-th root. It transforms under Γ through a certain character $\omega_n : \Gamma \longrightarrow \mathbb{F}_q^* \hookrightarrow C^*$ of precise order $q-1$ (*loc. cit.* 3.21, 3.22). Therefore, D_n^{q-1} (but no smaller power of D_n) is Γ-invariant, and it has the asserted divisor $\sharp(\mathcal{C})([0]-[\infty])$ on $\overline{M}_\Gamma$. Put finally

$$t := \gcd(q-1, \sharp(\mathcal{C})). \tag{5.8.1}$$

Then yet

$$\mathrm{div}(D_n^{(q-1)/t} = \frac{\sharp(\mathcal{C})}{t}([0]-[\infty])$$

is an integral divisor, whose class generates the subgroup U_t of order t in $\mathcal{C}$. A look at (4.7) shows that U_t is contained in the kernel of can_∞, with which it must agree in view of (5.7).

5.9 Theorem. *Let n be an irreducible monic polynomial of degree d in $A = \mathbb{F}_q[T]$, let $\Gamma = \Gamma_0(n)$ be the Hecke congruence subgroup, and t as given in* (5.8.1).
(i) *There is an exact sequence $0 \longrightarrow U_t \longrightarrow \mathcal{C} \stackrel{\mathrm{can}_\infty}{\longrightarrow} \phi_\infty$, where U_t is the unique subgroup of order t in $\mathcal{C} = \mathcal{C}(\Gamma)$.*
(ii) *The cokernel $\phi_\infty/\phi_\infty^{cusp}$ of can_∞ has order a multiple of t.*

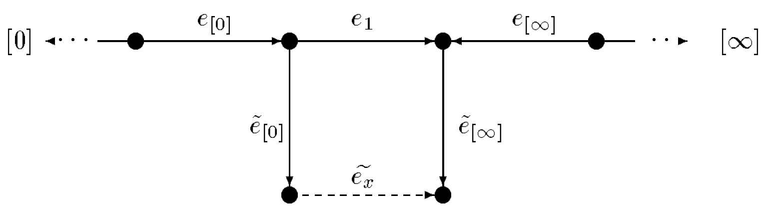
Proof. (i) has been shown. (ii) comes from (5.5), noting that $[\underline{H}_!^\perp : \underline{H}_!^\perp \cap r(\Theta_c')] = (q-1)/t$. $\square$

Pairs (q,d) where $t>1$ are for example (4,3), (7,3), (13,3) with $t=3$ and (3,4), (5,4) with $t=2$. In the final section, we work out an example with $(q,d) = (7,3)$.

6. An example.

We consider in detail the case where n is a prime of degree 3 in $A = \mathbb{F}_q[T]$. The graph $\Gamma \setminus \mathcal{T}$ looks ([4] 5.3, $\Gamma := \Gamma_0(n)$):

(6.1)



Here $\dashrightarrow$ stands for q edges $\tilde{e}_x$ indexed by $x \in \mathbb{F}_q$. The multiplicities $m(e)$ (see (5.4.4)) of all drawn edges and their inverses are 1 except for $\tilde{e}_{[0]}$ and $\tilde{e}_{[\infty]}$, which enter with multiplicity $q-1$ into the harmonicity condition w.r.t. their origins. Hence e.g.

$$(q-1)\varphi(\tilde{e}_{[\infty]}) - \varphi(e_1) - \varphi(e_{[\infty]}) = 0$$

for $\varphi \in \underline{H}$. The scalar product on $\underline{H}_! = \underline{H}_!(\mathcal{T},\mathbb{Z})^\Gamma$ is such that each pair $\{e, \overline{e}\}$ of inversely oriented edges contributes volume 1 except for $\{e_1, \overline{e}_1\}$, which has volume $q-1$. For each $x \in \mathbb{F}_q$, let φ_x be the unique element of $\underline{H}_!$ with

$$\varphi_x(\tilde{e}_{[\infty]}) = -1, \ \varphi_x(\tilde{e}_y) = \delta_{x,y} \quad (y \in \mathbb{F}_q).$$

Let further $\psi \in \underline{H}$ be such that

$$\psi(e_{[0]}) = 1 = \psi(e_1) = -\psi(e_{[\infty]})$$

and ψ vanishes off the line from $[0]$ to $[\infty]$. Next, let $\delta \in \underline{H}$ be defined as

$$\delta = \sum_{x \in \mathbb{F}_q} \varphi_x + (q^2 + q + 1)\psi.$$

Then, as is easily verified:

(6.2) (i) $\{\varphi_x \mid x \in \mathbb{F}_q\}$ is a basis of $\underline{H}_!$.

(ii) $\{\varphi_x \mid x \in \mathbb{F}_q\} \cup \{\psi\}$ is a basis of $\underline{H}$.

(iii) $\underline{H}_!^\perp = \mathbb{Z}\delta$

(iv) $r(\Theta_c') = \underline{H}_! + (q-1)\mathbb{Z}\psi$ (use (5.4)!)

(v) $\underline{H}_!^\perp \cap r(\Theta_c') = \frac{q-1}{t}\mathbb{Z}\delta$ $(t := \gcd(q-1, q^2+q+1))$

(vi) $P_\infty = (q-1)\mathbb{Z}\delta$ (see (4.3)).

Furthermore,

(6.3) (i) $\mathcal{C} = r(\Theta_c')/\underline{H}_! \oplus P_\infty \xrightarrow{\cong} \mathbb{Z}/(q^2+q+1)\mathbb{Z}$
$\varphi \longmapsto \varphi(e_{[0]})$

(in accordance with (5.6)) and

(ii) $\sharp(\phi_\infty) = q^2 + q + 1 = \sharp(\mathcal{C})$ (from calculating the determinant of $(.,.) : \underline{H}_! \times \underline{H}_! \longrightarrow \mathbb{Z}$), but $\mathrm{can}_\infty : \mathcal{C} \longrightarrow \phi_\infty$ has kernel and cokernel each of size t. (It is easy to show that in this case, ϕ_∞ is cyclic, too.)

(6.4) As is explained in [10], the splitting of the Jacobian $J := J_0(n)$ of $\overline{M}_\Gamma$ corresponds to the splitting of $\underline{H}_! \otimes \mathbb{Q}$ under the Hecke algebra, which can be calculated by the formulae in [4], or by the approach via modular symbols proposed in [23]. Let now, more specifically

(6.4.1) $q = 7$ and $n = T^3 - 2 \in \mathbb{F}_7[T]$, which gives $\sharp(\mathcal{C}) = 57$ and $t = \gcd(6, 57) = 3$. In that case, $\underline{H}_! \otimes \mathbb{Q}$ splits under the Hecke algebra into an irreducible piece of dimension 6 and the eigenspace generated by (see [4], table 10.3)

$$\varphi = \sum_{x \in \mathbb{F}_7} a_x \varphi_x \text{ with } (a_0, \ldots, a_6) = (4, 1, 1, -2, 1, -2, -2). \tag{6.4.2}$$

This means, there exists an elliptic curve E/K, uniquely determined up to isogeny, with good reduction outside of the two places ∞, (n) of $K = \mathbb{F}_7(T)$, multiplicative reduction at (n) and split multiplicative reduction at ∞, which has a "Weil uniformization" $\pi : \overline{M}_\Gamma \longrightarrow E$, and whose reduction at $(T-x)$ has $8 + a_x$ rational points over $A/(T-x) = \mathbb{F}_7$. We have

(6.4.3) $(\varphi, \varphi) = 39$, $m := \min\{(\varphi, \alpha) > 0 \mid \alpha \in \underline{H}_!\} = 3$, hence ([6] 3.19, 3.20) $\deg \pi = 39/3 = 13$ and $v_\infty(j_E) = -3$ for the j-invariant j_E of E, π supposed to be a "strong Weil uniformization". Comparing with [4] table 9.3, case 3a and performing the unramified quadratic twist to get split multiplicative reduction at ∞ yields the following equation for E:

$$Y^2 = X^3 + aX + b \tag{6.4.4}$$

with $a = -3T(T^3+2)$, $b = -2T^6+3T^3+1$. It can be shown by routine methods that (6.4.4) in fact yields the strong Weil curve in the given isogeny class, and that

$$E(K) = \{0, (3T^2, \pm 4(T^3-2))\} \cong \mathbb{Z}/3\mathbb{Z}. \tag{6.4.5}$$

(We should note here that the equation given in [23] p. 289, dealing with the same example, does not describe the isogeny factor E of J but its unramified quadratic twist. Hence some conclusions derived there must be slightly modified.)

Similar to (4.7), there is a map $\mathrm{can}_{\infty,E} : \mathcal{C}_E \longrightarrow \phi_{\infty,E}$ and a commutative diagram

$$\begin{array}{ccc} \mathcal{C} & \xrightarrow{\mathrm{can}_\infty} & \phi_\infty \\ \downarrow & & \downarrow \\ \mathcal{C}_E & \xrightarrow{\mathrm{can}_{\infty,E}} & \phi_{\infty,E}, \end{array} \tag{6.4.6}$$

where $\mathcal{C}_E$ is the image of the map $\mathcal{C} \longrightarrow E(K)$ derived from π and $\phi_{\infty,E}$ the group of connected components of E at ∞, isomorphic with $\mathbb{Z}/m\mathbb{Z} = \mathbb{Z}/3\mathbb{Z}$. Further, as results from the calculation of Hecke operators, $\mathcal{C} \longrightarrow E(K)$ is non-trivial, hence $\mathcal{C} \longrightarrow \mathcal{C}_E = E(K) \cong \mathbb{Z}/3\mathbb{Z}$, and E is the quotient of J corresponding to the *Eisenstein prime number* $l = 3$ ([15], [22]). Since, by (5.9), can_∞ kills the subgroup of order $t = 3$ in $\mathcal{C}$, (6.4.6) forces $\mathrm{can}_{\infty,E}$ to be trivial. In other words:

(6.4.7) The rational 3-division points (6.4.5) of E map to the connected component of the Néron model at ∞.

Of course, this is easy to see directly. An equivalent form of stating this fact is as follows: Let $f \in \Theta_c(\Gamma)$ be such that $r(f) = \delta$, and regard $\varphi \in \underline{H}_! \xleftarrow{\cong} \overline{\Gamma}$ as the class of some element of Γ. Then f^6 is a modular unit and, up to scaling, a 6-th root of Δ/Δ_n. Its third root f^2 belongs to $\Theta_c'(\Gamma)$ and transforms under Γ through a character $\chi = c_{f^2}$, and $\chi(\varphi)$ is a *non-trivial* third root of unity.

(6.5) The above example (and similar ones) suggests to refine the investigation (begun in [3] and, much more deeply, in [22]) of the Eisenstein ideal, the Eisenstein quotient of J etc., i.e., of data defined by means of the cuspidal divisor class group $\mathcal{C}(\Gamma)$, by taking into account the Hecke module $\phi_\infty(\Gamma)$ and the map $\mathrm{can}_\infty : \mathcal{C}(\Gamma) \longrightarrow \phi_\infty(\Gamma)$.

References

[1] J. Fresnel, M. van der Put: Géométrie analytique Rigide et Applications. Progr. Math. 18, Birkhäuser, Basel-Boston 1981.

[2] E.-U. Gekeler: Modulare Einheiten für Funktionenkörper. J. Reine Angew. Math. 348, 94-115 (1984).

[3] E.-U. Gekeler: Über Drinfeld'sche Modulkurven vom Hecke-Typ. Comp. Math. 57, 219-236 (1986).

[4] E.-U. Gekeler: Automorphe Formen über $\mathbb{F}_q(T)$ mit kleinem Führer. Abh. Math. Sem. Univ. Hamburg 55, 111-146 (1985).

[5] E.-U. Gekeler: Drinfeld Modular Curves. Lecture Notes Math. 1231, Springer 1986.

[6] E.-U. Gekeler: Analytical Construction of Weil Curves over Function Fields. J. Théorie des Nombres Bordeaux 7, 27-49 (1995).

[7] E.-U. Gekeler: On the Drinfeld Discriminant Function. Comp. Math. 106, 181-202 (1997).

[8] E.-U. Gekeler: A note on the finiteness of certain cuspidal divisor class groups, in preparation.

[9] E.-U. Gekeler, U. Nonnengardt: Fundamental Domains of some Arithmetic Groups over Function Fields. Int. J. Math. 6, 689-708 (1995).

[10] E.-U. Gekeler, M. Reversat: Jacobians of Drinfeld Modular Curves. J. Reine Angew. Math. 476, 27-93 (1996).

[11] L. Gerritzen, M. van der Put: Schottky Groups and Mumford Curves. Lecture Notes Math. 817, Springer 1980.

[12] D. Goss: Basic Structures of Function Field Arithmetic. Ergeb. Math. 35, Springer 1996.

[13] D. Kubert, S. Lang: Modular units. Grundlehren Math. Wiss. 244, Springer 1981.

[14] Y. Manin, V.G. Drinfeld: Periods of p-adic Schottky Groups. J. Reine Angew. Math. 262/263, 239-247 (1973).

[15] B. Mazur: Modular Curves and the Eisenstein Ideal. Publ. Math. IHES 47, 33-186 (1977).

[16] D. Mumford: An Analytic Construction of Degenerating Curves over Local Fields. Comp. Math. 24, 129-174 (1972).

[17] D. Mumford: An Analytic Construction of Degenerating Abelian Varieties over Complete Rings. Comp. Math. 24, 239-272 (1972).

[18] A. Scholl: Fourier Coefficients of Eisenstein Series on Non-Congruence Subgroups. Math. Proc. Camb. Phil. Soc. 99, 11-17 (1986).

[19] A. Schweizer: Zur Arithmetik der Drinfeld'schen Modulkurven $X_0(\mathfrak{n})$. Thesis, Saarbrücken 1996.

[20] J.-P. Serre: Trees. Springer 1980.

[21] G. Shimura: Introduction to the Arithmetic Theory of Automorphic Functions. Publ. Math. Soc. Japan, Tokyo-Princeton 1971.

[22] A. Tamagawa: The Eisenstein Quotient of the Jacobian Variety of a Drinfeld Modular Curve. Publ. RIMS, Kyoto Univ. 31, 203-246 (1995).

[23] J. Teitelbaum: Modular Symbols for $\mathbb{F}_q(T)$. Duke Math. J. 68, 271-295 (1992).

[24] M. van der Put: Les Fonctions thêta d'une Courbe de Mumford. Groupe d'Etude d'Analyse Ultramétrique 1981/82, Paris 1982.

Ernst-Ulrich Gekeler
FB 9 Mathematik
Universität des Saarlandes
Postfach 15 11 50
D-66041 Saarbrücken
Germany
gekeler@math.uni-sb.de

Stability of C^*-Algebras is Not a Stable Property

Mikael Rørdam

Received: November 28, 1997

Communicated by Joachim Cuntz

Abstract. We show that there exists a C^*-algebra B such that $M_2(B)$ is stable, but B is not stable. Hence stability of C^*-algebras is not a stable property. More generally, we find for each integer $n \geq 2$ a C^*-algebra B so that $M_n(B)$ is stable and $M_k(B)$ is not stable when $1 \leq k < n$. The C^*-algebras we exhibit have the additional properties that they are simple, nuclear and of stable rank one.

The construction is similar to Jesper Villadsen's construction in [7] of a simple C^*-algebra with perforation in its ordered K_0-group.

1991 Mathematics Subject Classification: 46L05, 46L35, 19K14

Keywords: Stable C^*-algebras, perforation in K_0, scaled ordered Abelian groups.

1 Introduction

A C^*-algebra A is said to be stable if $A \cong A \otimes \mathcal{K}$, where $\mathcal{K}$ is the C^*-algebra of compact operators on a separable, infinite dimensional Hilbert space. The problem of deciding which C^*-algebras are stable relates to structure problems of simple C^*-algebras. For example, as shown in [3, Proposition 5.2], if all non-unital hereditary sub-C^*-algebras of a given C^*-algebra A are stable, and if A is simple and not of type I, then A must be purely infinite. It was also remarked in [3, Proposition 5.1] that an AF-algebra is stable if and only if it admits no bounded (densely defined) traces, and it was asked if a similar characterization might hold in general. In more detail, is a C^*-algebra A stable if and only if A admits no bounded (quasi-)trace and no quotient of A is unital?

It is a consequence of the examples produced in this article that the answer to this question is no. Indeed, let A be a C^*-algebra such that $M_2(A)$ is stable and A is not stable. Then $M_2(A)$ admits no bounded (quasi-)trace, and no quotient of $M_2(A)$ is unital. This is easily seen to imply that A admits no bounded (quasi-)trace, and that no quotient of A is unital.

Jesper Villadsen gave in [7] the first examples of simple C^*-algebras whose ordered K_0-groups have perforation. As shown in Proposition 3.3, the examples constructed here must also have perforation in their K_0-group (at least when they admit an approximate unit consisting of projections). We shall in this article make extensive use of the techniques developed by Villadsen.

2 A preliminary result

Let A be a C^*-algebra and consider the set $\Gamma(A)$ consisting of those integers $n \geq 1$ where $M_n(A)$ is stable. The result below shows that this set must be either empty, $\mathbb{N}$, or equal to $\{n, n+1, n+2, \dots\}$ for some $n \geq 2$. Clearly, the empty set and $\mathbb{N}$ arise as $\Gamma(A)$ for appropriate C^*-algebras A. The main result of this article (Theorem 5.3) shows that the remaining sets are also realized.

Proposition 2.1 *Let A be a σ-unital C^*-algebra, let $n \geq 1$ be an integer, and suppose that $M_n(A)$ is stable. Then $M_{n+1}(A)$ is stable.*

Proof: By [3, Theorem 2.1 and Proposition 2.2] it suffices to show that one for all positive elements $a \in M_{n+1}(A)$ and all $\varepsilon > 0$ can find positive elements $b, c \in M_{n+1}(A)$ with $\|a-b\| \leq \varepsilon$, $\|bc\| \leq \varepsilon$, and $b \sim c$ (i.e. $b = x^*x$ and $c = xx^*$ for some $x \in M_{n+1}(A)$). To obtain this it suffices to find positive elements $e, f \in M_{n+1}(A)^+$ with $e \sim f$, $e \perp f$, and ea close to a. Indeed, if $e = x^*x$ and $f = xx^*$, then set $y = xa^{1/2}$, and note that y^*y is close to a and that $(yy^*)(y^*y)$ is small.

Now,

$$a = \begin{pmatrix} a_1 & z \\ z^* & a_2 \end{pmatrix},$$

where $a_1 \in M_n(A)^+$, $a_2 \in A^+$ and $z \in M_{n,1}(A)$. Let $\varepsilon > 0$, and let $\varphi_\varepsilon \colon \mathbb{R}^+ \to [0,1]$ be a continuous function which is zero on $[0, \varepsilon/2]$ and equal to 1 on $[\varepsilon, \infty)$. Set

$$e' = \begin{pmatrix} \varphi_\varepsilon(a_1) & 0 \\ 0 & \varphi_\varepsilon(a_2) \end{pmatrix}.$$

Then $e'a$ is close to to a if $\varepsilon > 0$ is small.

Since $M_n(A)$ is stable, we can find positive elements $e_1, f_1, f_2 \in M_n(A)$ and $e_2 \in A$ such that $e_1 \sim f_1$, $e_2 \sim f_2$ (in the sense that $e_2 = x^*x$ and $f_2 = xx^*$ for some $x \in M_{n,1}(A)$), e_1, f_1, f_2 are mutually orthogonal, e_1 is close to $\varphi_\varepsilon(a_1)$, and e_2 is close to $\varphi_\varepsilon(a_2)$. Set

$$e = \begin{pmatrix} e_1 & 0 \\ 0 & e_2 \end{pmatrix}, \qquad f = \begin{pmatrix} f_1 + f_2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then ea is close to a, $e \sim f$, and $e \perp f$ as desired. $\square$

3 Stability and the scale of K_0

We investigate in this section the connection between the scaled ordered group of a C^*-algebra and stability of matrix algebras over the C^*-algebra. Recall that if A is a C^*-algebra, then

$$K_0(A)^+ = \{[p]_0 \mid p \in P(A \otimes \mathcal{K})\} \subseteq K_0(A), \quad \Sigma(A) = \{[p]_0 \mid p \in P(A)\} \subseteq K_0(A)^+,$$

where $P(A \otimes \mathcal{K})$ and $P(A)$ denote the set of projections in $A \otimes \mathcal{K}$, respectively, A.

One can in some cases see from the triple $(K_0(A), K_0(A)^+, \Sigma(A))$ if A is stable. A C^*-algebra A is said to have the *cancellation property* if $p + r \sim q + r$ implies that $p \sim q$ for all projections $p, q, r \in A \otimes \mathcal{K}$ with $p \perp r$ and $q \perp r$. If A has the cancellation property, then $[p]_0 = [q]_0$ in $K_0(A)$ implies $p \sim q$ for all projections $p, q \in A \otimes \mathcal{K}$. Recall also that A has the cancellation property if A is of stable rank one (see [1, Proposition 6.5.1]).

Proposition 3.1 *Let A be a C^*-algebra with the cancellation property and with a countable approximate unit consisting of projections. Then A is stable if and only if $\Sigma(A) = K_0(A)^+$.*

Proof: The "only if" part is trivial. To show the "if" part, assume that $\Sigma(A) = K_0(A)^+$. By [3, Theorem 3.3] it suffices to show that for each projection $p \in A$ there exists a projection $q \in A$ with $p \sim q$ and $p \perp q$. Let a projection $p \in A$ be given. By the assumptions that A has an approximate unit consisting of projections, and $\Sigma(A) = K_0(A)^+$, there exist projections $e, f \in A$ such that $[e]_0 = 2[p]_0 = [p \oplus p]_0$, $e \leq f$ and $p \leq f$. Since A has the cancellation property, this implies that $e \sim p \oplus p$, which again implies that $e = e_1 + e_2$, where $e_1 \sim e_2 \sim p$. Now, $[f - p]_0 = [f - e_1]_0$, and so $p \sim e_2 \leq f - e_1 \sim f - p$. Hence p is equivalent to a subprojection q of $f - p$ as desired. $\Box$

Definition 3.2 *A triple (G, G^+, Σ) will be called a* scaled, ordered abelian group *if (G, G^+) is an ordered abelian group, and Σ is an upper directed, hereditary, full subset of G^+, ie.,*

(i) $\forall x_1, x_2 \in \Sigma \; \exists x \in \Sigma : x_1 \leq x, \; x_2 \leq x,$

(ii) $\forall x \in G^+ \; \forall y \in \Sigma : x \leq y \implies x \in \Sigma,$

(iii) $\forall x \in G^+ \; \exists y \in \Sigma \; \exists k \in \mathbb{N} : x \leq ky.$

Let (G, G^+) be an ordered abelian group, and let Σ_1 and Σ_2 be upper directed, hereditary, full subsets of G^+. Define $\Sigma_1 \hat{+} \Sigma_2$ to be the set of all elements $x \in G^+$ for which there exist $x_1 \in \Sigma_1$ and $x_2 \in \Sigma_2$ with $x \leq x_1 + x_2$. Observe that $\Sigma_1 \hat{+} \Sigma_2$ is an upper directed, hereditary, full subset of G^+. Denote the k-fold sum $\Sigma \hat{+} \Sigma \hat{+} \cdots \hat{+} \Sigma$ by $k \hat{\cdot} \Sigma$. Using that Σ is upper directed we see that $y \in k \hat{\cdot} \Sigma$ if and only if $0 \leq y \leq kx$ for some $x \in \Sigma$.

If A is a stably finite C^*-algebra with the cancellation property and with an approximate unit consisting of projections, then $(K_0(A), K_0(A)^+, \Sigma(A))$ is a scaled, ordered abelian group. If A has these properties, then

$$\big(K_0(M_k(A)), K_0(M_k(A))^+, \Sigma(M_k(A))\big) \cong (K_0(A), K_0(A)^+, k \hat{\cdot} \Sigma(A)). \tag{1}$$

Suppose that $n \geq 2$ and that (G, G^+, Σ) is a scaled, ordered Abelian group such that $(n-1) \hat{\cdot} \Sigma \neq G^+$ and $n \hat{\cdot} \Sigma = G^+$, and suppose that A is a C^*-algebra of stable rank one and with an approximate unit of projections such that $(K_0(A), K_0(A)^+, \Sigma(A)) \cong (G, G^+, \Sigma)$. Then it follows from Proposition 3.1 and (1) that $M_n(A)$ is stable and $M_k(A)$ is not stable for $1 \leq k < n$.

Recall that an ordered Abelian group (G, G^+) is called *weakly unperforated* if $ng \in G^+ \setminus \{0\}$ for some $n \in \mathbb{N}$ and some $g \in G$ implies $g \in G^+$.

PROPOSITION 3.3 *Let (G, G^+, Σ) be a weakly unperforated, scaled, ordered, Abelian group, and suppose that $n \hat{\cdot} \Sigma = G^+$ for some $n \in \mathbb{N}$. Then $\Sigma = G^+$.*

Proof: Let g be an element of G^+, and choose a non-zero element $u \in G^+$. Since $n \hat{\cdot} \Sigma = G^+$, there is an element $x \in \Sigma$ with $nx \geq ng + u$. Now, $n(x - g) \geq u > 0$, and this entails that $x - g \geq 0$, by the assumption that (G, G^+) is weakly unperforated. By the hereditary property of Σ we get that $g \in \Sigma$. Thus $\Sigma = G^+$. $\square$

We give below an explicit example of a scaled, ordered Abelian group (G, G^+, Σ) with $\Sigma \hat{+} \Sigma = G^+$ and $\Sigma \neq G^+$. Note that this ordered group necessarily must be perforated (by Proposition 3.3 above).

It is not known if every (countable) scaled ordered Abelian group is the scaled ordered Abelian group of a C^*-algebra — the problem here lies in realizing the given order structure, not in realizing the given scale. We can therefore not immediately conclude from the example below that there exists a non-stable C^*-algebra B where $M_2(B)$ is stable. Actually, it is not known (to the author) if the ordered Abelian group constructed below is the ordered K_0-group of any C^*-algebra.

EXAMPLE 3.4 Let $\mathbb{Z}_2$ denote the group $\mathbb{Z}/2\mathbb{Z}$, and let $\mathbb{Z}_2^{(\infty)}$ denote the group of all sequences $t = (t_j)_{j=1}^\infty$, with $t_j \in \mathbb{Z}_2$ and where $t_j \neq 0$ only for finitely many j. For each $t \in \mathbb{Z}_2^{(\infty)}$, let $d(t)$ be the number of elements in $\{j \in \mathbb{N} \mid t_j \neq 0\}$. Set

$$G_2 = \mathbb{Z} \oplus \mathbb{Z}_2^{(\infty)}, \qquad G_2^+ = \{(k, t) \mid d(t) \leq k\}, \qquad \Sigma_2 = \{(k, t) \mid d(t) = k\}.$$

Then (G_2, G_2^+, Σ_2) is a scaled, ordered Abelian group with $\Sigma_2 \neq G_2^+$ and $\Sigma_2 \hat{+} \Sigma_2 = G_2^+$. To see this, let $e_j \in \mathbb{Z}_2^{(\infty)}$ be the element which is a generator of $\mathbb{Z}_2$ at the jth coordinate and zero elsewhere, set $g_j = (1, e_j) \in G^+$, and set $h_j = g_1 + g_2 + \cdots + g_j$. Then

$$\Sigma_2 = \bigcup_{j=1}^{\infty} \{x \in G^+ \mid x \leq h_j\}. \tag{2}$$

The claims made about (G_2, G_2^+, Σ_2) are now easy to verify.

Notice that $\Sigma_2 + \Sigma_2 \neq \Sigma_2 \hat{+} \Sigma_2$, since for example $(3, e_1 + e_2) \notin \Sigma_2 + \Sigma_2$. This was pointed out to me by Jacob Hjelmborg, and it shows that the sum of two scales is not a scale in general. $\square$

EXAMPLE 3.5 Let $n \geq 2$ be an integer. Let $\mathbb{Z}_n^{(\infty)}$ be the Abelian group of all sequences $(t_j)_{j=1}^\infty$ with $t_j \in \mathbb{Z}_n$ $(= \mathbb{Z}/n\mathbb{Z})$, and $t_j \neq 0$ only for finitely many j. Let $e_j \in \mathbb{Z}_n^{(\infty)}$ be a generator of the jth copy of $\mathbb{Z}_n$. Then each $t \in \mathbb{Z}_n^{(\infty)}$ is a sum $t = \sum_{j=1}^\infty r_j e_j$ with $0 \leq r_j < n$ and where $r_j = 0$ for all but finitely many j. Set $d(t) = \sum_{j=1}^\infty r_j$, and set

$$G_n = \mathbb{Z} \oplus \mathbb{Z}_n^{(\infty)}, \qquad G_n^+ = \{(k, t) \mid d(t) \leq k\}, \qquad \Sigma_n = \bigcup_{j=1}^{\infty} \{x \in G^+ \mid x \leq h_j\},$$

where $g_j = (1, e_j)$ and $h_j = g_1 + g_2 + \cdots + g_j$. Then (G_n, G_n^+, Σ_n) is a scaled, ordered, Abelian group, $(n-1) \hat{\cdot} \Sigma_n \neq G_n^+$ and $n \hat{\cdot} \Sigma_n = G_n^+$. $\square$

Adopt the following (standard) notation. If $e \in M_n(A)$ and $f \in M_m(A)$ are projections, then let $e \oplus f$ denote the projection $\mathrm{diag}(e, f) \in M_{n+m}(A)$. Write $e \sim f$ if there is an element $v \in M_{m,n}(A)$ with $e = v^*v$ and $f = vv^*$, and write $e \precsim f$ if $e \sim f_0$ for some subprojection f_0 of f. Denote the k-fold direct sum $e \oplus e \oplus \cdots \oplus e$ by $e \otimes 1_k$. If A has the cancellation property (see the introduction to this section), and if $e, f \in A$ are projections, then $[e]_0 \leq [f]_0$ if and only if $e \precsim f$.

Proposition 3.6 *Let A be a C^*-algebra, let $n \geq 2$ be an integer, and suppose that A contains projections $e, p_1, p_2, p_3, \ldots$ that satisfy*

(i) *$e \otimes 1_n \sim p_j \otimes 1_n$ for all j,*

(ii) *e is not equivalent to a subprojection of $(p_1 \oplus p_2 \oplus \cdots \oplus p_j) \otimes 1_{n-1}$ for any j.*

Set $q_j = p_1 \oplus p_2 \oplus \cdots \oplus p_j$, and embed all matrix algebras over A coherently into $A \otimes \mathcal{K}$ so that q_j belongs to $A \otimes \mathcal{K}$ for all j. Set

$$B = \overline{\bigcup_{j=1}^{\infty} q_j (A \otimes \mathcal{K}) q_j}. \tag{3}$$

Then $M_k(B)$ is not stable for $1 \leq k < n$, but $M_n(B)$ is stable.

Let H be the subgroup of $K_0(B)$ generated by the K_0-classes of the projections $e, p_1, p_2, p_3, \ldots$. Assume that B has the cancellation property. Then

$$(H, H \cap K_0(B)^+, H \cap \Sigma(B)) \cong (G_n, G_n^+, \Sigma_n), \tag{4}$$

where the triple on the right hand-side is the scaled, ordered, Abelian group defined in Example 3.5.

Proof: Observe that

$$M_k(B) = \overline{\bigcup_{j=1}^{\infty} (q_j \otimes 1_k)(A \otimes \mathcal{K})(q_j \otimes 1_k)},$$

for each k, and that $\{q_j \otimes 1_k\}_{j=1}^{\infty}$ is an approximate unit for $M_k(B)$.

To show that $M_k(B)$ is not stable for $1 \leq k < n$ it suffices by Proposition 2.1 to show that $M_{n-1}(B)$ is not stable.

If $M_{n-1}(B)$ were stable, then there would exist a projection $q \in M_{n-1}(B)$ such that $q \sim p_1 \otimes 1_{n-1}$ and $q \perp p_1 \otimes 1_{n-1}$. (This is rather easy to see directly, and one can also obtain this from [3, Theorem 3.3].) Since $\{q_j \otimes 1_{n-1} - p_1 \otimes 1_{n-1}\}_{j=1}^{\infty}$ is an approximate unit for $(1 - p_1 \otimes 1_{n-1}) M_{n-1}(B)(1 - p_1 \otimes 1_{n-1})$, there is a j, so that q is equivalent to a subprojection of $q_j \otimes 1_{n-1} - p_1 \otimes 1_{n-1}$ $(= (p_2 \oplus p_3 \oplus \cdots \oplus p_j) \otimes 1_{n-1})$. By assumption (i),

$$\begin{aligned} e &\precsim e \otimes 1_n \sim p_1 \otimes 1_n \precsim (p_1 \otimes 1_{n-1}) \oplus (p_1 \otimes 1_{n-1}) \precsim (p_1 \otimes 1_{n-1}) \oplus q \\ &\precsim (p_1 \oplus p_2 \oplus \cdots \oplus p_j) \otimes 1_{n-1}, \end{aligned}$$

in contradiction with assumption (ii).

We proceed to show that $M_n(B)$ is stable. By (i), $q_j \otimes 1_n$ is equivalent to the direct sum of $e \otimes 1_n$ with itself j times. It follows quite easily from this that $M_n(B)$ is stable. We can also use [3, Theorem 3.3] to obtain this conclusion by showing that there for each projection p in $M_n(B)$ exists a projection q in $M_n(B)$ with $p \sim q$ and $p \perp q$. One can here reduce to the case where p is a subprojection of $q_j \otimes 1_n$ for some j, and the result then follows from the fact that $q_{2j} \otimes 1_n - q_j \otimes 1_n \sim q_j \otimes 1_n$.

Assume now that B has the cancellation property. To establish the isomorphism (4), note first that $n([p_j]_0 - [e]_0) = 0$ by (i). Retaining the notation from Example 3.5, we define a group homomorphism $\varphi\colon G_n \to H$ by $\varphi(1,0) = [e]_0$ and $\varphi(0,e_j) = [p_j]_0 - [e]_0$. φ is clearly surjective. For any $(k,t) \in G_n$ with $t = \sum_{j=1}^N r_j e_j$, $0 \le r_j < n$,

$$\varphi(k,t) = k[e]_0 + \sum_{j=1}^N r_j\big([p_j]_0 - [e]_0\big) = \big(k - d(t)\big)[e]_0 + \sum_{j=1}^N r_j[p_j]_0.$$

It follows that $\varphi(k,t) \ge 0$ if $(k,t) \ge 0$. Conversely, if (k,t) is not positive, then $k - d(t) \le -1$, and so

$$\varphi(k,t) = \big(k - d(t)\big)[e]_0 + \sum_{j=1}^N r_j[p_j]_0 \le (n-1)\big([p_1]_0 + [p_2]_0 + \cdots + [p_N]_0\big) - [e]_0.$$

By (ii) and the assumption that B has the cancellation property, the element on the right-hand side of this inequality is not positive. All in all we have shown that $\varphi(k,t) \ge 0$ if and only if $(k,t) \ge 0$. This entails that φ is injective and that $\varphi(G_n^+) = H \cap K_0(B)^+$.

Since $\{q_j\}_{j=1}^\infty$ is an approximate unit for B, an element $g \in K_0(B)$ lies in $\Sigma(B)$ if and only if $0 \le g \le [q_j]_0$ for some j. Notice that $\varphi(h_j) = [q_j]_0$. Hence $\varphi(k,t) \in \Sigma(B)$ if and only if $0 \le (k,t) \le h_j$ for some j, and this shows that $\varphi(\Sigma_n) = H \cap \Sigma(B)$. □

REMARK 3.7 Corollary 4.2 and Proposition 5.2 contain for each prime number n examples of C^*-algebras with projections $e, p_1, p_2, p_3, \dots$ satisfying (i) and (ii) of Proposition 3.6. The C^*-algebras in Proposition 5.2 have the cancellation property (being of stable rank one).

REMARK 3.8 One can replace condition (i) in Proposition 3.6 by a weaker condition such as for example $e \precsim p_j \otimes 1_n$ for all j, and still obtain that the C^*-algebra B defined in (3) has the property that $M_k(B)$ is not stable for $1 \le k < n$ and $M_n(B)$ is stable. However, with this weaker condition one would not have a description of the scaled ordered group as in (4).

4 THE COMMUTATIVE CASE

We realize for each positive prime number n projections $e, p_1, p_2, p_3, \dots$ satisfying conditions (i) and (ii) of Proposition 3.6, with respect to that n, inside a C^*-algebra which is stably isomorphic to a commutative C^*-algebra. At the same time, Lemma 4.1 below, is a key ingredient in Section 5.

If $\pi\colon X_1 \to X_2$ is a continuous function, then π^* will denote the map from the cohomology groups of X_2 to the cohomology groups of X_1, and the same symbol will be used to denote the map from vector bundles over X_2 to vector bundles over X_1. By naturality of the Euler class, $\mathrm{e}(\pi^*(\xi)) = \pi^*(\mathrm{e}(\xi))$ for all complex vector bundles ξ over Y.

The proof of Lemma 4.1 below is almost identical to the proof of [6, Theorem 3.4]. The statements of Lemma 4.1 and of [6, Theorem 3.4] are, however, quite different. Therefore, and for the convenience of the reader, we include a proof of Lemma 4.1.

Let $\mathbb{D}$ denote the unit disk in the complex plane. Consider for each integer $n \geq 2$ the equivalence relation $\sim$ on $\mathbb{D}$ given by: $z \sim w$ if $z = w$ or if $|z| = |w| = 1$ and $z^n = w^n$. Put $Y_n = \mathbb{D}/\sim$.

Lemma 4.1 *Let n be a positive prime number, and put $X = Y_n^{n-1}$. There exists a complex line bundle ω over X with the following properties. Let m be a positive integer, let $\pi_1, \pi_2, \dots, \pi_m \colon X^m \to X$ be the coordinate maps, and set*

$$\xi_k^{(m)} = \pi_1^*(\omega) \oplus \pi_2^*(\omega) \oplus \cdots \oplus \pi_k^*(\omega), \qquad 1 \leq k \leq m,$$

which is a complex vector bundle over X^m of dimension k. Let θ_d denote the trivial complex vector bundle (over X or X^m) of (complex) dimension d. Then

(i) *$n\omega \cong \theta_n$,*

(ii) *if $(n-1)\xi_k^{(m)} \oplus \theta_{d_1} \cong \eta \oplus \theta_{d_2}$ for some complex vector bundle η over X^m, and some positive integers d_1 and d_2, then $d_1 \geq d_2$, and*

(iii) *$\omega \oplus \eta \cong \theta_n$ for some $(n-1)$-dimensional complex vector bundle η over X.*

Proof: Recall that $H^2(Y_n; \mathbb{Z}) \cong \mathbb{Z}/n\mathbb{Z}$. There is a complex line bundle ζ over Y_n with non-trivial Euler class $\mathrm{e}(\zeta) \in H^2(Y_n; \mathbb{Z})$, and with $n\zeta \cong \theta_n$. Let $\nu_1, \nu_2, \dots, \nu_{n-1} \colon X = Y_n^{n-1} \to Y_n$ be the coordinate projections, and set

$$\omega = \nu_1^*(\zeta) \otimes \nu_2^*(\zeta) \otimes \cdots \otimes \nu_{n-1}^*(\zeta).$$

Then ω is a complex line bundle over X, and successive applications of the isomorphism $n\zeta \cong \theta_n = n\theta_1$, yield $n\omega \cong \theta_n$. Hence (i) holds, and (iii) is a trivial consequence of (i).

To prove claim (ii) we first show that the Euler class, $\mathrm{e}((n-1)\xi_k^{(m)})$, is non-zero. The Euler class of ω is given by

$$\mathrm{e}(\omega) = \sum_{j=1}^{n-1} \nu_j^*(\mathrm{e}(\zeta)), \tag{5}$$

cf. [4, Proposition V.3.10]. By the product formula for the Euler class, cf. [4, Proposition V.3.10],

$$\mathrm{e}((n-1)\xi_k^{(m)}) = \prod_{j=1}^{k} \pi_j^*(\mathrm{e}(\omega)^{n-1}). \tag{6}$$

Since $\mathrm{e}(\zeta)^2 \in H^4(Y_n; \mathbb{Z})$ and $H^4(Y_n; \mathbb{Z}) = 0$, it follows from (5) and (6) that

$$\mathrm{e}(\omega)^{n-1} = (n-1)! \prod_{i=1}^{n-1} \nu_i^*(\mathrm{e}(\zeta)).$$

Let $\rho_1, \rho_2, \dots, \rho_k \colon X^k \to X$ and $\pi \colon X^m \to X^k$ be the projections maps. Then $\pi_j = \rho_j \circ \pi$, and $\pi^* \colon H^{2k}(X^k; \mathbb{Z}) \to H^{2k}(X^m; \mathbb{Z})$ is an injection. The map

$$\mu \colon H^2(Y_n; \mathbb{Z}) \otimes H^2(Y_n; \mathbb{Z}) \otimes \cdots \otimes H^2(Y_n; \mathbb{Z}) \to H^{2k(n-1)}(X^k; \mathbb{Z})$$

given by

$$\mu(x_{1,1} \otimes x_{1,2} \otimes \cdots \otimes x_{k,n-1}) = \prod_{j=1}^{k} \prod_{i=1}^{n-1} (\rho_j^* \circ \nu_i^*)(x_{i,j}),$$

is injective by the Künneth formula. Now,

$$\begin{aligned} \mathrm{e}((n-1)\xi_k^{(m)}) &= \prod_{j=1}^{k} \pi_j^*(\mathrm{e}(\omega)^{n-1}) \\ &= \prod_{j=1}^{k} \pi_j^*\big((n-1)! \prod_{i=1}^{n-1} \nu_i^*(\mathrm{e}(\zeta))\big) \\ &= (n-1)!^k \pi^*\big(\prod_{j=1}^{k} \prod_{i=1}^{n-1} (\rho_j^* \circ \nu_i^*)(\mathrm{e}(\zeta)\big) \\ &= (\pi^* \circ \mu)\big((n-1)!^k \mathrm{e}(\zeta) \otimes \mathrm{e}(\zeta) \otimes \cdots \otimes \mathrm{e}(\zeta)\big). \end{aligned}$$

The element $\mathrm{e}(\zeta) \otimes \mathrm{e}(\zeta) \otimes \cdots \otimes \mathrm{e}(\zeta)$ has order n in $H^2(Y_n; \mathbb{Z}) \otimes H^2(Y_n; \mathbb{Z}) \otimes \cdots \otimes H^2(Y_n; \mathbb{Z})$. Because n is assumed to be prime, and because $\pi^* \circ \mu$ is injective, we get that $\mathrm{e}((n-1)\xi_k^{(m)}) \neq 0$.

Assume (ii) were false. Then $(n-1)\xi_k^{(m)} \oplus \theta_{d_1} \cong \eta \oplus \theta_{d_2}$ for some η and some positive integers $d_1 < d_2$. Hence $(n-1)\xi_k^{(m)}$ would be stably isomorphic to $\eta \oplus \theta_{d_2 - d_1}$. The Euler class is invariant under stable isomorphism, and the Euler class of a trivial bundle (of dimension ≥ 1) is zero, and so by the product formula we get $\mathrm{e}((n-1)\xi_k^{(m)}) = 0$, a contradiction. $\square$

George Elliott pointed out to me that one obtains the following corollary from Lemma 4.1:

COROLLARY 4.2 *Let n be a positive prime number, let Z be the infinite Cartesian product of Y_n with itself. Then there exist projections $e, p_1, p_2, p_3, \dots$ in $M_n(C(Z))$ satisfying*

(i) *$e \otimes 1_n \sim p_j \otimes 1_n$ for all j,*

(ii) *e is not equivalent to a subprojection of $(p_1 \oplus p_2 \oplus \cdots \oplus p_j) \otimes 1_{n-1}$ for any $j \geq 1$.*

Proof: Let ω be the complex line bundle over $X = Y_n^{n-1}$ from Lemma 4.1 and use Lemma 4.1 (iii) to find a projection $p \in C(X, M_n(\mathbb{C})) = M_n(C(X))$ that corresponds to ω. Identify Z with $\prod_{j=1}^{\infty} X$, and let $\pi_j \colon Z \to X$, $j \in \mathbb{N}$, be the coordinate maps. Put $p_j = p \circ \pi_j \in C(Z, M_n(\mathbb{C})) = M_n(C(Z))$, and let $e \in M_n(C(Z))$ be a one-dimensional constant projection. It follows from Lemma 4.1 (i) that $p_j \otimes 1_n \sim e \otimes 1_n$ for all j. To see (ii), view $M_n(C(Z))$ as the inductive limit,

$$M_n(C(X)) \to M_n(C(X^2)) \to M_n(C(X^3)) \to \cdots \to M_n(C(Z)),$$

so that $e, p_1, p_2, \ldots, p_j \in M_n(C(X^j))$. Then, by Lemma 4.1 (ii), for each k and for each $m \geq k$, e is not equivalent to a subprojection of $(p_1 \oplus p_2 \oplus \cdots \oplus p_k) \otimes 1_{n-1}$ in (a matrix algebra over) $M_n(C(X^m))$. This implies that (ii) holds. □

Combining Corollary 4.2 with Proposition 3.6 we get for each prime number n a hereditary sub-C^*-algebra B of $C(Z) \otimes \mathcal{K}$ such that $M_k(B)$ is not stable for $1 \leq k < n$, and $M_n(B)$ is stable. Proceeding as in the proof of Theorem 5.3 one can find such examples B for all integers $n \geq 2$.

5 THE SIMPLE CASE

We use an inductive limit construction, like the one Villadsen used in [7], to obtain projections as in Proposition 3.6 inside a *simple* C^*-algebra.

Fix a positive prime number n. Let $\{k_j\}_{j=1}^{\infty}$ be a sequence of positive integers chosen large enough so that

$$\sum_{j=1}^{\infty} (1 - \prod_{i=j}^{\infty} \frac{k_i}{1 + k_i}) < \frac{1}{n-1}. \tag{7}$$

Define inductively another sequence of integers $\{m_j\}_{j=1}^{\infty}$ by $m_1 = 1$ and $m_{j+1} = m_j(k_j + 1)$.

Let $Y_n = \mathbb{D}/\sim$ be as defined in Section 4, and put $X = Y_n^{n-1}$. Define inductively a sequence of spaces $\{X_j\}_{j=1}^{\infty}$ by setting $X_1 = X$ and $X_{j+1} = X_j^{k_j} \times X^{m_{j+1}}$. Set

$$A_j = M_{2^{n-1}m_j}(C(X_j)) = C(X_j, M_{2^{n-1}m_j}(\mathbb{C})).$$

Choose $x_j \in X_j$ appropriately (in a way which will be made precise later), and define *-homomorphisms $\varphi_j : A_j \to A_{j+1}$ by

$$\varphi_j(f)(x) = \operatorname{diag}((f \circ \pi_1^j)(x), (f \circ \pi_2^j)(x), \ldots, (f \circ \pi_{k_j}^j)(x), f(x_j)), \quad x \in X_{j+1}, \quad f \in A_j,$$

where $\pi_1^j, \pi_2^j, \ldots, \pi_{k_j}^j : X_{j+1} = X_j^{k_j} \times X^{m_{j+1}} \to X_j$ are the projections from the first factor of X_{j+1}.

Let $(A, \mu_j : A_j \to A)$ be the inductive limit of the sequence

$$A_1 \xrightarrow{\varphi_1} A_2 \xrightarrow{\varphi_2} A_3 \xrightarrow{\varphi_3} \cdots.$$

It will be convenient to have an expression for the composed connecting maps $\varphi_{i,j} : A_j \to A_i$ for $i > j$. For this purpose set

$$k_{i,j} = \prod_{n=j}^{i-1} k_n, \qquad l_{i,j} = \prod_{n=j}^{i-1}(k_n + 1) - \prod_{n=j}^{i-1} k_n, \qquad m_{i,j} = \sum_{n=j+1}^{i} m_n k_{i,n}, \tag{8}$$

(with the convention that $k_{i,i} = 1$). Then $X_i = X_j^{k_{i,j}} \times X^{m_{i,j}}$, and the composed connecting maps are up to unitary equivalence given by

$$\begin{aligned} &\varphi_{i,j}(f)(x) \\ &\quad = \operatorname{diag}\big((f \circ \pi_1^{i,j})(x), (f \circ \pi_2^{i,j})(x), \ldots, (f \circ \pi_{k_{i,j}}^{i,j})(x), f(x_1^{i,j}), f(x_2^{i,j}), \ldots, f(x_{l_{i,j}}^{i,j})\big). \end{aligned}$$

The maps $\pi_1^{i,j}, \pi_2^{i,j}, \dots, \pi_{k_{i,j}}^{i,j} \colon X_i = X_j^{k_{i,j}} \times X^{m_{i,j}} \to X_j$ are here the projections onto the first $k_{i,j}$ coordinates of X_i, the set

$$X_j^i := \{x_1^{i,j}, x_2^{i,j}, \dots, x_{l_{i,j}}^{i,j}\} \subseteq X_j$$

is for $i \geq j+2$ equal to $X_j^{i-1} \cup \{\pi_1^{i,j}(x_i), \pi_2^{i,j}(x_i), \dots, \pi_{k_{i,j}}^{i,j}(x_i)\}$, where each element of the first set is repeated $k_i + 1$ times, and $X_j^{j+1} = \{x_j\}$.

Choose the points $x_j \in X_j$ such that $\bigcup_{r=j+1}^{\infty} X_j^r$ is dense in X_j for each $j \in \mathbb{N}$. Since each X_j^i is finite and since no X_j has isolated points this will entail that $\bigcup_{r=i}^{\infty} X_j^r$ is dense in X_j for each $j \in \mathbb{N}$ and for every $i > j$.

By [2, Proposition 1] and [7, Proposition 10] we get:

Proposition 5.1 *The C^*-algebra A is simple and has stable rank one.*

With the C^*-algebra A and the prime number n as above, we have:

Proposition 5.2 *There exist projections $e, p_1, p_2, p_3, \dots$ in A so that*

(i) *$p_j \otimes 1_n \sim e \otimes 1_n$ for all $j \geq 1$, and*

(ii) *e is not equivalent to a subprojection of $(p_1 \oplus p_2 \oplus \cdots \oplus p_j) \otimes 1_{n-1}$ for any $j \geq 1$.*

Proof: By Lemma 4.1 (iii) there exists a projection $q \in A_1 = M_{2^{n-1}}(C(X))$ which corresponds to the complex line bundle ω. Let $\rho_1, \rho_2, \dots, \rho_{m_j} \colon X_j = X_{j-1}^{k_{j-1}} \times X^{m_j} \to X$ be coordinate projections corresponding to the last factor of X_j. Set $q_1 = q$, set

$$q_j = \operatorname{diag}(q \circ \rho_1, q \circ \rho_2, \dots, q \circ \rho_{m_j}) \in A_j,$$

for $j \geq 2$, and set $p_j = \mu_j(q_j) \in A$. Let $e_1 \in A_1$ be a constant projection of dimension 1, so that e_1 corresponds to the trivial complex line bundle θ_1, and set $e = \mu_1(e_1) \in A$.

It follows from Lemma 4.1 (i) that $q \otimes 1_n \sim e_1 \otimes 1_n$. This implies that $q_j \otimes 1_n$ is equivalent to a constant projection. Since $\varphi_{j,1}(e_1) \otimes 1_n$ is a constant projection (in $M_n(A_j)$) of the same dimension as $q_j \otimes 1_n$, we find that $q_j \otimes 1_n \sim \varphi_{j,1}(e_1) \otimes 1_n$ in $M_n(A_j)$. Hence

$$p_j \otimes 1_n = \mu_j(q_j \otimes 1_n) \sim \mu_j(\varphi_{j,1}(e_1) \otimes 1_n) = e \otimes 1_n$$

in $M_n(A)$.

For $i \geq j$, put

$$f_{i,j} = \varphi_{i,1}(q_1) \oplus \varphi_{i,2}(q_2) \oplus \cdots \oplus \varphi_{i,j}(q_j).$$

Then $p_1 \oplus p_2 \oplus \cdots \oplus p_j = \mu_i(f_{i,j})$, and $f_{i,j} = \varphi_{i,j}(f_{j,j})$. Observe that $X_j = X^{d_j}$, where $d_1 = 1$ and $d_{j+1} = n_j k_j + m_{j+1}$. By inspection of the formula for the composed connecting maps $\varphi_{j,l}$, we find that the projection $f_{j,j}$ corresponds to the vector bundle $\xi_{d_j}^{(d_j)} \oplus \theta_{c_j}$, where $c_j = \sum_{r=1}^{j} m_r l_{j,r}$, cf. (8). From this we get that the projection $f_{i,j}$ corresponds to the vector bundle $\xi_{a_{i,j}}^{(d_i)} \oplus \theta_{b_{i,j}}$ over X_i, where $a_{i,j} = k_{i,j} d_j$ and $b_{i,j} = \sum_{r=1}^{j} m_r l_{i,r}$, possibly after a permutation of the coordinates of X_i.

The trivial projection $\varphi_{i,1}(e_1)$ has dimension m_i and corresponds therefore to the trivial vector bundle θ_{m_i}. Now,

$$\begin{aligned}\frac{1}{m_i}b_{i,j} &= \frac{1}{m_i}\sum_{r=1}^{j} m_r l_{i,r} \\ &= \frac{1}{m_i}\sum_{r=1}^{j}\prod_{s=1}^{r-1}(1+k_s)\Big(\prod_{s=r}^{i-1}(1+k_s)-\prod_{s=r}^{i-1}k_s\Big) \\ &= \sum_{r=1}^{j}\Big(1-\prod_{s=r}^{i-1}\frac{k_s}{1+k_s}\Big) \\ &\le \sum_{r=1}^{\infty}\Big(1-\prod_{s=r}^{\infty}\frac{k_s}{1+k_s}\Big) < \frac{1}{n-1},\end{aligned}$$

where the last inequality follows from (5). This shows that $(n-1)b_{i,j} < m_i$. By Lemma 4.1 (ii), there exists no vector bundle η over X_i such that

$$\eta \oplus \theta_{m_i} \cong (n-1)\xi^{(d_i)}_{a_{i,j}} \oplus \theta_{(n-1)b_{i,j}} \; (= (n-1)(\xi^{(d_i)}_{a_{i,j}} \oplus \theta_{b_{i,j}})),$$

or, equivalently, $\varphi_{i,1}(e_1)$ is not equivalent to a subprojection of $f_{i,j} \otimes 1_{n-1}$. Since this holds for all $i > j$, e is not equivalent to a subprojection of $(p_1 \oplus p_2 \oplus \cdots \oplus p_j) \otimes 1_{n-1}$, and this completes the proof. □

THEOREM 5.3 *For each integer $n \ge 2$ there exists a C^*-algebra B such that $M_n(B)$ is stable, and $M_k(B)$ is not stable for $1 \le k < n$. Moreover, B can be chosen to be simple, nuclear, with stable rank one and with an approximate unit consisting of projections.*

Proof: Consider first the case where n is prime. Let B be the C^*-algebra defined in display (3) in Proposition 3.6 corresponding to the C^*-algebra A and to the projections $e, p_1, p_2, p_3, \ldots$ found in Proposition 5.2. Then B is a hereditary subalgebra of $A \otimes \mathcal{K}$, and since A is simple, nuclear and has stable rank one, it follows that B also has these properties (see [5, Theorem 3.3] for the last claim). The sequence $\{q_j\}_{j=1}^{\infty}$ is an approximate unit for B. By Proposition 3.6, $M_k(B)$ is not stable for $1 \le k < n$ and $M_n(B)$ is stable.

Suppose now that $n \ge 2$ is an arbitrary integer. Observe that all integers $\ge (n-1)^2$ belong to the set

$$\bigcup_{m=1}^{\infty} ((n-1)m, nm].$$

Choose a prime number $p \ge (n-1)^2$. Then there exists an integer $m \ge 1$ so that $(n-1)m < p \le nm$. By the first part of the proof there exists a C^*-algebra D with $M_p(D)$ stable and $M_k(D)$ not stable for $1 \le k < p$. Set $B = M_m(D)$. Then B is simple, nuclear, and has stable rank one and an approximate unit consisting of projections because D has these properties. Moreover, $M_k(B) = M_{km}(D)$, and so, by Proposition 2.1, $M_k(B)$ is stable if and only if $km \ge p$, which, by the choice of p and m, happens if and only if $k \ge n$. □

References

[1] B. Blackadar, *K-theory for operator algebras*, M. S. R. I. Monographs, vol. 5, Springer Verlag, Berlin and New York, 1986.

[2] M. Dădărlat, G. Nagy, A. Némethi, and C. Pasnicu, *Reduction of topological stable rank in inductive limits of C^*-algebras*, Pacific J. Math. 153 (1992), 267–276.

[3] J. Hjelmborg and M. Rørdam, *On stability of C^*-algebras*, J. Funct. Anal., to appear.

[4] M. Karoubi, *K-theory*, Grundlehren der mathematisches Wissenschaften, no. 226, Springer Verlag, 1978.

[5] M.A. Rieffel, *Dimension and stable rank in the K-theory of C^*-algebras*, Proc. London Math. Soc. 46 (1983), no. (3), 301–333.

[6] M. Rørdam and J. Villadsen, *On the ordered K_0-group of universal free product C^*-algebras*, preprint.

[7] J. Villadsen, *Simple C^*-algebras with perforation*, J. Funct. Anal., to appear.

Mikael Rørdam
Department of Mathematics
University of Copenhagen
Universitetsparken 5
2100 Copenhagen Ø, Denmark
rordam@math.ku.dk