On the unit group of some multiquadratic number fields

Article in Pacific Journal of Mathematics · March 2007
DOI: 10.2140/pjm.2007.230.27

CITATION
1

READS
28

3 authors, including:

Elliot Benjamin
Capella University
37 PUBLICATIONS 176 CITATIONS
SEE PROFILE

Chip Snyder
University of Maine
36 PUBLICATIONS 267 CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

History of number theory View project

All content following this page was uploaded by Elliot Benjamin on 05 September 2014.
The user has requested enhancement of the downloaded file.
PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

Founded in 1951 by

EDITORS

V. S. Varadarajan (Managing Editor)
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
pacific@math.ucla.edu

Vyjayanthi Chari
Department of Mathematics
University of California
Riverside, CA 92521-0135
chari@math.ucr.edu

Darren Long
Department of Mathematics
University of California
Santa Barbara, CA 93106-3080
long@math.ucsb.edu

Sorin Popa
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
popa@math.ucla.edu

Robert Finn
Department of Mathematics
Stanford University
Stanford, CA 94305-2125
finn@math.stanford.edu

Jiang-Hua Lu
Department of Mathematics
University of Hong Kong
Pokfulam Rd., Hong Kong
jlu@maths.hku.hk

Jie Qing
Department of Mathematics
University of California
Santa Cruz, CA 95064
qing@cats.ucsc.edu

Kefeng Liu
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
liu@math.ucla.edu

Alexander Merkurjev
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
merkurev@math.ucla.edu

Jonathan Rogawski
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
jor@math.ucla.edu

PRODUCTION

pacific@math.berkeley.edu

Paulo Ney de Souza, Production Manager
Silvio Levy, Senior Production Editor
Alexandru Scorpan, Production Editor

SUPPORTING INSTITUTIONS

ACADEMIA SINICA, TAIPEI
UNIVERSIDAD DE LOS ANDES
UNIV. OF CALIF., SANTA CRUZ
CALIFORNIA INST. OF TECHNOLOGY
UNIV. OF ARIZONA
UNIV. OF HAWAII
INST. DE MATEMÁTICA PURA E APLICADA
UNIV. OF BRITISH COLUMBIA
UNIV. OF MONTANA
KEIO UNIVERSITY
UNIV. OF CALIFORNIA, BERKELEY
UNIV. OF NEVADA, RENO
MATH. SCIENCES RESEARCH INSTITUTE
UNIV. OF CALIFORNIA, DAVIS
UNIV. OF OREGON
NEW MEXICO STATE UNIV.
UNIV. OF CALIFORNIA, IRVINE
UNIV. OF SOUTHERN CALIFORNIA
OREGON STATE UNIV.
UNIV. OF CALIFORNIA, LOS ANGELES
UNIV. OF UTAH
PEKING UNIVERSITY
UNIV. OF CALIFORNIA, RIVERSIDE
UNIV. OF WASHINGTON
STANFORD UNIVERSITY
UNIV. OF CALIFORNIA, SAN DIEGO
WASHINGTON STATE UNIVERSITY
UNIV. OF CALIF., SANTA BARBARA

These supporting institutions contribute to the cost of publication of this Journal, but they are not owners or publishers and have no responsibility for its contents or policies.

See inside back cover or www.pjmath.org for submission instructions.

Regular subscription rate for 2007: $425.00 a year (10 issues). Special rate: $212.50 a year to individual members of supporting institutions. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 969 Evans Hall, Berkeley, CA 94720-3840 is published monthly except July and August. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS
at the University of California, Berkeley 94720-3840
A NON-PROFILE CORPORATION
Typeset in LATEX
Copyright ©2007 by Pacific Journal of Mathematics
We define and study the Tutte polynomial of a hyperplane arrangement. We introduce a method for computing the Tutte polynomial by solving a related enumerative problem. As a consequence, we obtain new formulas for the generating functions enumerating alternating trees, labelled trees, semiorders and Dyck paths.

1. Introduction

Much work has been devoted in recent years to studying hyperplane arrangements and, in particular, their characteristic polynomials. The polynomial $\chi_{\mathcal{A}}(q)$ is a very powerful invariant of the arrangement $\mathcal{A}$; it arises very naturally in many different contexts. Two of the many important and beautiful results about the characteristic polynomial of an arrangement are the following.

**Theorem 1.1** [Zaslavsky 1975]. Let $\mathcal{A}$ be a hyperplane arrangement in $\mathbb{R}^n$. The number of regions into which $\mathcal{A}$ dissects $\mathbb{R}^n$ is equal to $(-1)^n \chi_{\mathcal{A}}(-1)$. The number of regions which are relatively bounded is equal to $(-1)^n \chi_{\mathcal{A}}(1)$.

**Theorem 1.2** [Orlik and Solomon 1980]. Let $\mathcal{A}$ be a hyperplane arrangement in $\mathbb{C}^n$, and let $M_{\mathcal{A}} = \mathbb{C}^n - \bigcup_{H \in \mathcal{A}} H$ be its complement. Then the Poincaré polynomial of the cohomology ring of $M_{\mathcal{A}}$ is given by:

$$
\sum_{k \geq 0} \text{rank } H^k(M_{\mathcal{A}}, \mathbb{Z}) q^k = (-q)^n \chi_{\mathcal{A}}(-1/q).
$$

Several authors have worked on computing the characteristic polynomials of specific hyperplane arrangements. This work has led to some very nice enumerative results; see for example [Athanasiadis 1996; Postnikov and Stanley 2000].

Somewhat surprisingly, nothing has been said about the Tutte polynomial of a hyperplane arrangement. Graphs and matroids have a Tutte polynomial associated with them, which generalizes the characteristic polynomial and arises naturally...
in numerous enumerative problems in both areas. Many interesting invariants of graphs and matroids can be computed immediately from this polynomial.

The present paper aims to define and investigate the Tutte polynomial of a hyperplane arrangement. This paper is devoted to purely enumerative questions, particularly the computation of Tutte polynomials of specific arrangements. We addressed the matroid-theoretic aspects of this investigation in [Ardila 2004].

In Section 2 we introduce the basic notions of hyperplane arrangements that we will need. In Section 3 we define the Tutte polynomial of a hyperplane arrangement, and we present a finite field method for computing it. This is done in terms of the coboundary polynomial, a simple transformation of the Tutte polynomial. We recover several known results about the characteristic and Tutte polynomials of graphs and representable matroids, and derive other consequences of this method. In Section 4 we compute the Tutte polynomials of Coxeter arrangements, threshold arrangements, and generic deformations of the braid arrangement. In Section 5 we focus on a large family of deformations of the braid arrangement, where the computation of Tutte polynomials is related to the enumeration of classical combinatorial objects. As a consequence, we obtain several purely enumerative results about alternating trees, parking functions, semiorders, and Dyck paths.

2. Hyperplane arrangements

We recall some of the basic concepts of hyperplane arrangements. For a more thorough introduction, see [Orlik and Terao 1992] or [Stanley 2004].

Given a field $k$ and a positive integer $n$, an affine hyperplane in $k^n$ is an $(n-1)$-dimensional affine subspace of $k^n$. If we fix a system of coordinates $x_1, \ldots, x_n$ on $k^n$, a hyperplane can be seen as the set of points that satisfy a certain equation $c_1 x_1 + \cdots + c_n x_n = c$, where $c_1, \ldots, c_n, c \in k$ and not all $c_i$s are equal to 0. A hyperplane arrangement $\mathcal{A}$ in $k^n$ is a finite set of affine hyperplanes of $k^n$. We will refer to hyperplane arrangements simply as arrangements. We will mostly be interested in arrangements in $\mathbb{R}^n$, but we will find it useful to work over other fields as well.

We will say that an arrangement $\mathcal{A}$ is central if the hyperplanes in $\mathcal{A}$ have a nonempty intersection. (Sometimes we will call an arrangement affine to emphasize that it does not need to be central.) Similarly, we will say that a subset (or subarrangement) $\mathcal{B} \subseteq \mathcal{A}$ of hyperplanes is central if the hyperplanes in $\mathcal{B}$ have a nonempty intersection.

The rank function $r_{\mathcal{A}}$ is defined for each central subset $\mathcal{B}$ by $r_{\mathcal{A}}(\mathcal{B}) = n - \dim \bigcap \mathcal{B}$. This function can be extended to a function $r_{\mathcal{A}} : 2^{\mathcal{A}} \to \mathbb{N}$, by defining the rank of a noncentral subset $\mathcal{B}$ to be the largest rank of a central subset of $\mathcal{B}$. The rank of $\mathcal{A}$ is $r_{\mathcal{A}}(\mathcal{A})$, and it is denoted $r_{\mathcal{A}}$. 
Alternatively, if the hyperplane $H$ has defining equation $c_1 x_1 + \cdots + c_n x_n = c$, associate its normal vector $v = (c_1, \ldots, c_n)$ to it. Then define $r_{\mathcal{A}}(\{H_1, \ldots, H_k\})$ to be the dimension of the span of the corresponding vectors $v_1, \ldots, v_k$ in $\mathbb{R}^n$. It is easy to see that these two definitions of the rank function agree. In particular, this means that the resulting function $r_{\mathcal{A}} : 2^\mathcal{A} \to \mathbb{N}$ is the rank function of a matroid.

We will usually omit the subscripts when the underlying arrangement is clear, and simply write $r(\mathcal{B})$ and $r$ for $r_{\mathcal{A}}(\mathcal{B})$ and $r_{\mathcal{A}}$, respectively.

(Note that there is another natural way to extend $r_{\mathcal{A}}$ to the rank function of a matroid; see [Ardila 2004].)

The rank function gives us natural definitions of the usual concepts of matroid theory, such as independent sets, bases, flats, and circuits, in the context of hyperplane arrangements. All of this is done more naturally in the broader context of semimatroids in [Ardila 2004].

To each hyperplane arrangement $\mathcal{A}$ we assign a partially ordered set, called the intersection poset of $\mathcal{A}$ and denoted $L_{\mathcal{A}}$. It consists of the nonempty intersections $H_{i_1} \cap \cdots \cap H_{i_k}$, ordered by reverse inclusion. This poset is graded, with rank function $r(H_{i_1} \cap \cdots \cap H_{i_k}) = r_{\mathcal{A}}(\{H_{i_1}, \ldots, H_{i_k}\})$, and a unique minimal element $\emptyset = \mathbb{R}^n$. We will call two arrangements $\mathcal{A}_1$ and $\mathcal{A}_2$ combinatorially isomorphic or simply isomorphic, and write $\mathcal{A}_1 \cong \mathcal{A}_2$, if $L_{\mathcal{A}_1} \cong L_{\mathcal{A}_2}$. Isomorphic arrangements may be defined over different fields.

The characteristic polynomial of $\mathcal{A}$ is

$$\chi_{\mathcal{A}}(q) = \sum_{x \in L_{\mathcal{A}}} \mu(\emptyset, x) q^{n-r(x)}.$$ 

where $\mu$ denotes the Möbius function of $L_{\mathcal{A}}$ [Stanley 1997, Section 3.7].

Let $\mathcal{A}$ be an arrangement and let $H$ be a hyperplane in $\mathcal{A}$. The arrangement $\mathcal{A} - \{H\}$ (or simply $\mathcal{A} - H$), obtained by removing $H$ from the arrangement, is called the deletion of $H$ in $\mathcal{A}$. It is an arrangement in $\mathbb{R}^n$. The arrangement $\mathcal{A}/H = \{H' \cap H \mid H' \in \mathcal{A} - H, H' \cap H \neq \emptyset\}$, consisting of the intersections of the other hyperplanes with $H$, is called the contraction of $H$ in $\mathcal{A}$. It is an arrangement in $H$.

However, some technical difficulties can arise. In a hyperplane arrangement $\mathcal{A}$, contracting a hyperplane $H$ may give us repeated hyperplanes $H_1$ and $H_2$ in the arrangement $\mathcal{A}/H$. Say we want to contract $H_1$ in $\mathcal{A}/H$. In passing to the contraction $(\mathcal{A}/H)/H_1$, the hyperplane $H_2$ of $\mathcal{A}/H$ becomes the “hyperplane” $H_2 \cap H_1 = H_1$ in the “arrangement” $(\mathcal{A}/H)/H_1$. But this is not a hyperplane in $H_1$.

Therefore, the class of hyperplane arrangements, as defined, is not closed under deletion and contraction. This is problematic when we want to mirror matroid-theoretic results in this context. There is an artificial solution to this problem: we can consider multisets $\{H_1, \ldots, H_k\}$ of subspaces of a vector space $V$, where each
$H_i$ has dimension $\dim V - 1$ or $\dim V$. In other words, we allow repeated hyperplanes, and we allow the full space $V$ to be regarded as a “hyperplane”, mirroring a loop of a matroid. This class of objects is closed under deletion and contraction, but it is somewhat awkward to work with. A better solution is to think of arrangements as members of the class of semimatroids: a class that is also closed under deletion and contraction, and is more natural matroid-theoretically. We develop this point of view in [Ardila 2004]. However, such issues will be irrelevant in this paper, which focuses on purely enumerative aspects of arrangements.

3. Computing the Tutte polynomial

Athanasiadis [1996] introduced a powerful method for computing the characteristic polynomial of a subspace arrangement, based on ideas of Crapo and Rota [1970]. He reduced the computation of characteristic polynomials to an enumeration problem in a vector space over a finite field. He used this method to compute explicitly the characteristic polynomial of several families of hyperplane arrangements, obtaining very nice enumerative results. As should be expected, this method only works when the equations defining the hyperplanes of the arrangement have integer (or rational) coefficients. Such an arrangement will be called a $\mathbb{Z}$-arrangement.

Reiner discovered an elegant interpretation for the Tutte polynomial $T_M$ of a representable matroid $M$ (Equation (3) in [1999]), and asked whether it could be used to compute $T_M$ for some nontrivial families of matroids. Compared to all the work that has been done on computing characteristic polynomials explicitly, virtually nothing is known about computing Tutte polynomials.

Our Theorem 3.3 below gives a new method for computing Tutte polynomials of hyperplane arrangements. Our approach does not use Reiner’s result; it is closer to Athanasiadis’ method. After proving the theorem, we present some of its consequences, using it in Section 4 to compute explicitly the Tutte polynomials of several families of arrangements.

The Tutte and coboundary polynomials.

**Definition 3.1.** The Tutte polynomial of a hyperplane arrangement $\mathcal{A}$ is

$$T_{\mathcal{A}}(x, y) = \sum_{\mathcal{B} \subseteq \mathcal{A}} (x - 1)^{r-\mathcal{B}} (y - 1)^{|\mathcal{B}| - r(\mathcal{B})},$$

where the sum is over all central subsets $\mathcal{B} \subseteq \mathcal{A}$.

If $\mathcal{A}$ is central and $M(\mathcal{A})$ is its associated matroid, this definition coincides with the usual definition of the Tutte polynomial of the matroid $M(\mathcal{A})$.

It will be useful for us to consider a simple transformation of the Tutte polynomial, first considered by Crapo [1969] in the context of matroids.
Definition 3.2. The coboundary polynomial $\chi_\mathcal{A}(q, t)$ of an arrangement $\mathcal{A}$ is

$$(3-2) \quad \chi_\mathcal{A}(q, t) = \sum_{\emptyset \subseteq B \subseteq \mathcal{A}} q^{r - r(B)} (t - 1)^{\lvert B \rvert}.$$  

It is easy to check that

$$\chi_\mathcal{A}(q, t) = (t - 1)^r T_\mathcal{A} \left( \frac{q + t - 1}{t - 1}, t \right)$$

and

$$T_\mathcal{A}(x, y) = \frac{1}{(y - 1)^r} \chi_\mathcal{A}((x - 1)(y - 1), y).$$

Therefore, computing the coboundary polynomial of an arrangement is essentially equivalent to computing its Tutte polynomial. The results in this paper can be presented more elegantly in terms of the coboundary polynomial.

Also, recall Whitney’s theorem [Stanley 2004, Theorem 2.4], which states that

$$\chi_\mathcal{A}(q) = \sum_{\emptyset \subseteq B \subseteq \mathcal{A}} (-1)^{\lvert B \rvert} q^{n - r(B)}.$$  

This allows us to express the characteristic polynomial in terms of the coboundary polynomial:

$$\chi_\mathcal{A}(q) = q^{n-r} \chi_\mathcal{A}(q, 0).$$

The finite field method. Let $\mathcal{A}$ be a $\mathbb{Z}$-arrangement in $\mathbb{R}^n$, and let $q$ be a prime power. The arrangement $\mathcal{A}$ induces an arrangement $\mathcal{A}_q$ in the vector space $\mathbb{F}_q^n$. If we consider the equations defining the hyperplanes of $\mathcal{A}$, and regard them as equations over $\mathbb{F}_q$, they define the hyperplanes of $\mathcal{A}_q$.

Say that $\mathcal{A}$ reduces correctly over $\mathbb{F}_q$ if the arrangements $\mathcal{A}$ and $\mathcal{A}_q$ are isomorphic. This does not always happen; sometimes the hyperplanes of $\mathcal{A}$ do not even become hyperplanes in $\mathcal{A}_q$. For example, the hyperplane $2x + 2y = 1$ in $\mathbb{R}^2$ becomes the empty “hyperplane” $0 = 1$ in $\mathbb{F}_2^2$. Sometimes independence is not preserved. For example, the independent hyperplanes $2x + y = 0$ and $y = 0$ in $\mathbb{R}^2$ become the same hyperplane in $\mathbb{F}_2^2$.

However, if $q$ is a power of a large enough prime, $\mathcal{A}$ will reduce correctly over $\mathbb{F}_q$. To have $\mathcal{A} \cong \mathcal{A}_q$, we need central and independent subarrangements to be preserved. Cramer’s rule lets us rephrase these conditions, in terms of certain determinants (formed by the coefficients of the hyperplanes in $\mathcal{A}$) being zero or nonzero. If $q$ is a power of a prime which is larger than all these determinants, we will guarantee that $\mathcal{A}$ reduces correctly over $\mathbb{F}_q$. 

Theorem 3.3. Let \( \mathcal{A} \) be a \( \mathbb{Z} \)-arrangement in \( \mathbb{R}^n \). Let \( q \) be a power of a large enough prime, and let \( \mathcal{A}_q \) be the induced arrangement in \( \mathbb{F}_q^n \). Then

\[
q^{n-r} \chi_{\mathcal{A}}(q, t) = \sum_{p \in \mathbb{F}_q^n} t^{h(p)}
\]

where \( h(p) \) denotes the number of hyperplanes of \( \mathcal{A}_q \) that \( p \) lies on.

Proof: Let \( q \) be a power of a large enough prime, so that \( \mathcal{A} \) reduces correctly over \( \mathbb{F}_q \). For each \( \mathcal{B} \subseteq \mathcal{A} \), let \( \mathcal{B}_q \) be the subarrangement of \( \mathcal{A}_q \) induced by it. For each \( p \in \mathbb{F}_q^n \), let \( H(p) \) be the set of hyperplanes of \( \mathcal{A}_q \) that \( p \) lies on. From (3-2) we have

\[
q^{n-r} \chi_{\mathcal{A}}(q, t) = \sum_{\mathcal{B} \subseteq \mathcal{A} \text{ central}} q^{n-r(\mathcal{B})}(t-1)^{|\mathcal{B}|} = \sum_{\mathcal{B} \subseteq \mathcal{A} \text{ central}} q^{\dim \mathcal{B} - \mathcal{B}_q}(t-1)^{|\mathcal{B}_q|} = \sum_{\mathcal{B}_q \subseteq \mathcal{A}_q \text{ central}} \sum_{p \in \mathcal{B}_q} (t-1)^{|\mathcal{B}_q|} = \sum_{p \in \mathbb{F}_q^n, \mathcal{B}_q \subseteq H(p)} (t-1)^{|\mathcal{B}_q|} = \sum_{p \in \mathbb{F}_q^n} (1 + (t-1))^{h(p)},
\]

as desired. \( \square \)

In principle, Theorem 3.3 only computes \( \chi_{\mathcal{A}}(q, t) \) when \( q \) is a power of a large enough prime. In practice, however, when we compute the right-hand side of (3-3) for large prime powers \( q \), we will get a polynomial function in \( q \) and \( t \). Since the left-hand side is also a polynomial, these two polynomials must be equal.

Theorem 3.3 reduces the computation of coboundary polynomials (and hence Tutte polynomials) to enumerating points in the finite vector space \( \mathbb{F}_q^n \), according to a certain statistic. This method can be extremely useful when the hyperplanes of the arrangement are defined by simple equations. We illustrate this in Section 4.

We remark that Theorem 3.3 was also obtained by Welsh and Whittle [1999, Theorem 7.1]. Also, since the characteristic polynomial of \( \mathcal{A} \) is given by \( \chi_{\mathcal{A}}(q) = q^{n-r} \chi_{\mathcal{A}}(q, 0) \), the special case \( t = 0 \) is the following result:

Theorem 3.4 [Athanasiadis 1996, Theorem 2.2]. If \( \mathcal{A} \) and \( q \) are as in Theorem 3.3, then \( \chi_{\mathcal{A}}(q) \) is the number of points in \( \mathbb{F}_q^n \) which are not on any of the hyperplanes of \( \mathcal{A}_q \); that is,

\[
\chi_{\mathcal{A}}(q) = |\mathbb{F}_q^n - \mathcal{A}_q|.
\]

Special cases and applications. We now show how the finite field method of Theorem 3.3 can be used to give straightforward proofs of four known facts (Theorems 3.5, 3.6, 3.7, 3.8) and an apparently new result (Theorem 3.9) on Tutte polynomials.
Colorings of graphs. Given a graph $G$ on $[n]$, we associate to it an arrangement $\mathcal{A}_G$ in $\mathbb{R}^n$. It consists of the hyperplanes $x_i = x_j$, for all $1 \leq i < j \leq n$ such that $ij$ is an edge in the graph $G$. It is easy to see that the matroid of $\mathcal{A}_G$ is the cycle matroid of $G$, so $T_{\mathcal{A}_G}(x, y)$ coincides with the (graph-theoretic) Tutte polynomial $T_G(x, y)$. We can define the coboundary polynomial for a graph as we did for arrangements, and then $\overline{\chi}_G(q, t) = \overline{\chi}_{\mathcal{A}_G}(q, t)$ also.

We shall now interpret Theorem 3.3 in this framework. It is easy to see that the rank of $\mathcal{A}_G$ is equal to $n - c$, where $c$ is the number of connected components of $G$. Therefore the left-hand side of (3-3) is $q^c \overline{\chi}_G(q, t)$ in this case.

To interpret the right-hand side, notice that each point $p \in \mathbb{R}^n$ corresponds to a $q$-coloring of the vertices of $G$. The point $p = (p_1, \ldots, p_n)$ will correspond to the coloring $\kappa_p$ of $G$ which assigns color $p_i$ to vertex $i$. A hyperplane $x_i = x_j$ contains $p$ when $p_i = p_j$. This happens precisely when edge $ij$ is monochromatic in $\kappa_p$; that is, when its two ends have the same color. Therefore, applying Theorem 3.3 to the arrangement $\mathcal{A}_G$, we recover the following known result:

**Theorem 3.5** [Brylawski and Oxley 1992, Proposition 6.3.26]. Let $G$ be a graph with $c$ connected components. Then

$$q^c \overline{\chi}_G(q, t) = \sum_{\kappa \text{colorings}} t^{\text{mono}(\kappa)}$$

where $\text{mono}(\kappa)$ is the number of monochromatic edges in $\kappa$.

Linear codes. Given positive integers $n \geq r$, an $[n, r]$ linear code $C$ over $\mathbb{F}_q$ is an $r$-dimensional subspace of $\mathbb{F}_q^n$. A generator matrix for $C$ is an $r \times n$ matrix $U$ over $\mathbb{F}_q$, the rows of which form a basis for $C$. It is not difficult to see that the isomorphism class of the matroid on the columns of $U$ depends only on $C$. We shall denote the corresponding matroid $M_C$.

The elements of $C$ are called codewords. The weight $w(v)$ of a codeword is the cardinality of its support; that is, the number of nonzero coordinates of $v$. The codeweight polynomial of $C$ is

$$A(C, q, t) = \sum_{v \in C} t^{w(v)}.$$  

The translation of Theorem 3.3 to this setting is the following.

**Theorem 3.6** [Greene 1976]. For any linear code $C$ over $\mathbb{F}_q$,

$$A(C, q, t) = t^n \overline{\chi}_{M_C}(q, 1/t).$$

**Proof.** Let $\mathcal{A}_C$ be the central arrangement corresponding to the columns of $U$. (We can call it $\mathcal{A}_C$ because, as stated above, its isomorphism class depends only on $C$.)
This is a rank $r$ arrangement in $\mathbb{F}_q^r$ such that $\overline{\mathcal{F}}_{\mathcal{M}_c}(q, \frac{1}{r}) = \overline{\mathcal{F}}_{\mathcal{R}_c}(q, \frac{1}{r})$. Comparing (3-4) with Theorem 3.3, it remains to prove that

$$\sum_{v \in C} t^{w(v)} = \sum_{p \in \mathbb{F}_q^r} t^{n-h(p)}.$$ 

To do this, consider the bijection $\phi : \mathbb{F}_q^r \to C$ determined by right multiplication by $U$. If $u_1, \ldots, u_r$ are the row vectors of $U$, then $\phi$ sends $p = (p_1, \ldots, p_r) \in \mathbb{F}_q^r$ to the codeword $v_p = p_1 u_1 + \cdots + p_r u_r \in C$. For $1 \leq i \leq n$, $p$ lies on the hyperplane determined by the $i$th column of $U$ if and only if the $i$th coordinate of $v_p$ is equal to zero. Therefore $h(p) = n - w(v_p)$. This completes the proof. \qed

**Deletion-contraction.** The point of view of Theorem 3.3 can be used to give a simple enumerative proof of the deletion-contraction formula for the Tutte polynomial of an arrangement. Once again, this formula is more natural in the context of semimatroids, as shown in [Ardila 2004]. For the moment, leaving matroid-theoretical issues aside, we only wish to present a special case of it as an application.

**Theorem 3.7.** Let $\mathcal{A}$ be a hyperplane arrangement, and let $H$ be a hyperplane in $\mathcal{A}$ such that $r_{\mathcal{A}}(\mathcal{A} - H) = r_{\mathcal{A}}$. Then $T_{\mathcal{A}}(x, y) = T_{\mathcal{A} - H}(x, y) + T_{\mathcal{A}/H}(x, y)$.

**Proof.** Because there will be several arrangements involved, let $h(\mathcal{B}, p)$ denote the number of hyperplanes in $\mathcal{B}_q$ that $p$ lies on. Then

$$q^{n-r} \overline{\mathcal{F}_{\mathcal{A}}}(q, t) = \sum_{p \in \mathbb{F}_q^n} t^{h(\mathcal{A}, p)} = \sum_{p \in \mathbb{F}_q^n - H} t^{h(\mathcal{A}, p)} + \sum_{p \in H} t^{h(\mathcal{A}, p)}$$

$$= \sum_{p \in \mathbb{F}_q^n - H} t^{h(\mathcal{A} - H, p)} + \sum_{p \in H} t^{h(\mathcal{A} - H, p)+1}$$

$$= \sum_{p \in \mathbb{F}_q^n - H} t^{h(\mathcal{A} - H, p)} + (t - 1) \sum_{p \in H} t^{h(\mathcal{A} - H, p)}$$

$$= q^{n-r} \overline{\mathcal{F}_{\mathcal{A} - H}}(q, t) + (t - 1) q^{(n-1)-(r-1)} \overline{\mathcal{F}_{\mathcal{A}/H}}(q, t).$$

We conclude that $\overline{\mathcal{F}_{\mathcal{A}}}(q, t) = \overline{\mathcal{F}_{\mathcal{A} - H}}(q, t) + (t - 1) \overline{\mathcal{F}_{\mathcal{A}/H}}(q, t)$, which is equivalent to the deletion-contraction formula for Tutte polynomials. \qed

**A Möbius formula.** The finite field method, when combined with the Möbius inversion formula for posets, naturally gives an alternative formula for the coboundary polynomial. This formula, in the context of matroids, is due to Crapo [1968].

**Theorem 3.8.** For an arrangement $\mathcal{A}$ and an affine subspace $x$ in the intersection poset $L_{\mathcal{A}}$, let $h(x)$ be the number of hyperplanes of $\mathcal{A}$ containing $x$. Then

$$\overline{\mathcal{F}_{\mathcal{A}}}(q, t) = \sum_{x \leq y \text{ in } L_{\mathcal{A}}} \mu(x, y) q^{t-r(y)} t^{h(x)}.$$
Proof. Consider the arrangement $\mathcal{A}$ restricted to $\mathbb{F}_q^n$, where $q$ is a power of a large enough prime, so that $\mathcal{A}$ reduces correctly over $\mathbb{F}_q$. Given $x \in L_{\mathcal{A}_q}$, let $P(x)$ be the set of points in $\mathbb{F}_q^n$ which are contained in $x$, and are not contained in any $y$ such that $y > x$ in $L_{\mathcal{A}_q}$. Then the set $x$ is partitioned by the sets $P(y)$ for $y \geq x$, so we have

$$q^{\dim x} = |x| = \sum_{y \geq x} |P(y)|.$$  

By the Möbius inversion formula [Stanley 1997, Proposition 3.7.1] we have

$$|P(x)| = \sum_{y \geq x} \mu(x, y) q^{\dim y}.$$  

Now, from Theorem 3.3 we know that

$$q^{n-r} \chi_{\mathcal{A}_q}(q, t) = \sum_{x \in L_{\mathcal{A}_q}} \sum_{p \in P(x)} t^h(p) = \sum_{x \in L_{\mathcal{A}_q}} |P(x)| t^h(x) = \sum_{x \leq y \text{ in } L_{\mathcal{A}_q}} \mu(x, y) q^{n-r(y)} t^h(x),$$

as desired. \qed

A probabilistic interpretation. In matroid reliability and percolation problems, one starts with a fixed matroid $M$. Each element of the ground set of $M$ has a certain probability of being deleted, independently of the other elements. One then asks for the probability that the retained elements satisfy a certain property. See [Brylawski and Oxley 1992, Section 6.3.E] for more on this subject. The following theorem is similar in spirit to these results, and it may be applied to the analogous questions concerning hyperplane arrangements.

**Theorem 3.9.** Let $\mathcal{A}$ be an arrangement and let $0 \leq t \leq 1$ be a real number. Let $\mathcal{B}$ be a random subarrangement of $\mathcal{A}$, obtained by independently removing each hyperplane from $\mathcal{A}$ with probability $t$. Then the expected characteristic polynomial $\chi_{\mathcal{B}_q}(q)$ of $\mathcal{B}$ is $q^{n-t} \chi_{\mathcal{A}_q}(q, t)$.

Proof. We have

$$E[\chi_{\mathcal{B}_q}(q)] = \sum_{\emptyset \subseteq \mathcal{B} \subseteq \mathcal{A}} \Pr[\mathcal{B} = \emptyset] \chi_{\emptyset}(q) = \sum_{\emptyset \subseteq \mathcal{B} \subseteq \mathcal{A}} \Pr[\mathcal{B} = \emptyset] |\mathbb{F}_q^n - \mathcal{B}_q|$$

$$= \sum_{p \in \mathbb{F}_q^n} \sum_{\emptyset \subseteq \mathcal{B} \subseteq \mathcal{A}} \Pr[\mathcal{B} = \emptyset] = \sum_{p \in \mathbb{F}_q^n} \Pr[\mathcal{B}_q = \emptyset],$$

where in the second step we have used Theorem 3.4.

Recall that $H(p)$ denotes the set of hyperplanes in $\mathcal{A}_q$ containing $p$. Then

$$E[\chi_{\mathcal{B}_q}(q)] = \sum_{p \in \mathbb{F}_q^n} \Pr[\mathcal{B}_q \cap H(p) = \emptyset] = \sum_{p \in \mathbb{F}_q^n} t^h(p),$$

which is precisely what we wanted to show. \qed
4. Computing coboundary polynomials

In this section we use Theorem 3.3 to compute the coboundary polynomials of several families of arrangements. As remarked on page 4, this is essentially the same as computing their Tutte polynomials.

**Coxeter arrangements.** To illustrate how our finite field method works, we start by presenting some simple examples.

Let $\Phi$ be an irreducible crystallographic root system in $\mathbb{R}^n$, and let $W$ be its associated Weyl group. The Coxeter arrangement of type $W$ consists of the hyperplanes $(\alpha, x) = 0$ for each $\alpha \in \Phi^+$, with the standard inner product on $\mathbb{R}^n$. See [Björner and Brenti 2005] or [Humphreys 1990] for an introduction to root systems and Weyl groups, and [Björner et al. 1993, Section 2.3] or [Orlik and Terao 1992, Chapter 6] for more information on Coxeter arrangements.

In this section we compute the coboundary polynomials of the Coxeter arrangements of type $A_n$, $B_n$, and $D_n$. (The arrangement of type $C_n$ is the same as the arrangement of type $B_n$.) The best way to state our results is to compute the exponential generating function for the coboundary polynomials of each family.

The following three theorems have never been stated explicitly in the literature in this form. Theorem 4.1 is equivalent to a result of Tutte [1954], who computed the Tutte polynomial of the complete graph. It is also an immediate consequence of a more general theorem of Stanley [1998a, (15)]. Theorems 4.2 and 4.3 are implicit in the work of Zaslavsky [1995].

**Theorem 4.1.** Let $A_n$ be the Coxeter arrangement (known as the braid arrangement) of type $A_{n-1}$ in $\mathbb{R}^n$, consisting of the hyperplanes $x_i = x_j$ for $1 \leq i < j \leq n$. We have

$$1 + q \sum_{n \geq 1} \chi_{A_n}(q, t) \frac{x^n}{n!} = \left( \sum_{n \geq 0} t^n \frac{x^n}{n!} \right)^q.$$ 

**Proof.** For $n \geq 1$ we have $q \chi_{A_n}(q, t) = \sum_{p \in F_q} t^{h(p)}$ for all powers of a large enough prime $q$, according to Theorem 3.3. For each $p \in F_q^n$, if we let $A_k = \{ i \in [n] \mid p_i = k \}$ for $0 \leq k \leq q - 1$, then $h(p) = \binom{|A_0|}{2} + \cdots + \binom{|A_{q-1}|}{2}$. Thus

$$q \chi_{A_n}(q, t) = \sum_{A_0 \cup \cdots \cup A_{q-1} = [n]} t^{\binom{|A_0|}{2} + \cdots + \binom{|A_{q-1}|}{2}}$$

summing over all weak ordered $q$-partitions of $[n]$; that is, ordered lists of $q$ pairwise disjoint, possibly empty sets whose union is $[n]$. The compositional formula for exponential generating functions [Bergeron et al. 1998; Stanley 1999, Proposition 5.1.3] implies the desired result. \[\square\]
Theorem 4.2. Let $\mathfrak{B}_n$ be the Coxeter arrangement of type $B_n$ in $\mathbb{R}^n$, consisting of the hyperplanes $x_i = x_j$ and $x_i + x_j = 0$ for $1 \leq i < j \leq n$, and the hyperplanes $x_i = 0$ for $1 \leq i \leq n$. We have
\[
\sum_{n \geq 0} \mathcal{X}_{\mathfrak{B}_n}(q, t) \frac{x^n}{n!} = \left( \sum_{n \geq 0} \frac{2^n t^n x^n}{n!} \right)^{q-1} \left( \sum_{n \geq 0} t^n x^n \right)^{\frac{q-1}{2}}.
\]

Proof. Let $q$ be a power of a large enough prime, and let $s = \frac{q-1}{2}$. Now for each $p \in \mathbb{F}_q$, if we let $B_k = \{ i \in [n] \mid p_i = k \}$ or $p_i = q - k$ for $0 \leq k \leq s$, we have that $h(p) = |B_0|^2 + (\binom{|B_1|}{2}) + \cdots + (\binom{|B_s|}{2})$. Also, given a weak ordered partition $(B_0, \ldots, B_s)$ of $[n]$, there are $2^{|B_1| + \cdots + |B_s|}$ points of $p$ which correspond to it: for each $i \in B_k$ with $k \neq 0$, we get to choose whether $p_i$ is equal to $k$ or to $q - k$. Therefore
\[
q \mathcal{X}_{\mathfrak{B}_n}(q, t) = \sum_{B_0 \cup \cdots \cup B_s = [n]} t^{|B_0|^2} \left( 2^{|B_1|} t^{|B_1|} \right) \cdots \left( 2^{|B_s|} t^{|B_s|} \right),
\]
and the compositional formula for exponential generating functions implies the theorem. \hfill \square

The proof of the next theorem is very similar to that of Theorem 4.2.

Theorem 4.3. Let $\mathfrak{D}_n$ be the Coxeter arrangement of type $D_n$ in $\mathbb{R}^n$, consisting of the hyperplanes $x_i = x_j$ and $x_i + x_j = 0$ for $1 \leq i < j \leq n$. We have
\[
\sum_{n \geq 0} \mathcal{X}_{\mathfrak{D}_n}(q, t) \frac{x^n}{n!} = \left( \sum_{n \geq 0} \frac{2^n t^n x^n}{n!} \right)^{q-1} \left( \sum_{n \geq 0} t^n (n-1) x^n \right)^{\frac{q-1}{2}}.
\]

Setting $t = 0$ in Theorems 4.1, 4.2 and 4.3, it is easy to recover the formulas for the characteristic polynomials of the above arrangements:
\[
\chi_{\mathfrak{S}_n}(q) = q(q-1)(q-2) \cdots (q-n+1),
\]
\[
\chi_{\mathfrak{A}_n}(q) = (q-1)(q-3) \cdots (q-2n+1),
\]
\[
\chi_{\mathfrak{W}_n}(q) = (q-1)(q-3) \cdots (q-2n+3)(q-n+1),
\]
which are well known; see for example [Stanley 2004].

Two more examples.

Theorem 4.4. Let $\mathcal{A}_n$ be a generic deformation of the arrangement $\mathcal{A}_n$, consisting of the hyperplanes $x_i - x_j = a_{ij}$ ($1 \leq i < j \leq n$), where the $a_{ij}$ are generic real
numbers\(^1\). For \(n \geq 1\),
\[
q \overline{\chi}_{\mathcal{B}^n}(q, t) = \sum_{F \text{ forest on } [n]} q^{n-e(F)}(t-1)^{e(F)},
\]
where \(e(F)\) denotes the number of edges of \(F\). Also,
\[
1 + q \sum_{n \geq 1} \overline{\chi}_{\mathcal{B}^n}(q, t) \frac{x^n}{n!} = \exp\left(\frac{q}{t-1} \sum_{n \geq 0} n^{n-2} \frac{x^n(t-1)^n}{n!}\right).
\]

**Proof.** It is possible to prove Theorem 4.4 using our finite field method, as we did in the previous section. However, it will be easier to proceed directly from (3-2), the definition of the coboundary polynomial.

To each subarrangement \(\mathcal{B}\) of \(\mathcal{B}^n\) we can assign a graph \(G_{\mathcal{B}}\) on the vertex set \([n]\), by letting edge \(ij\) be in \(G_{\mathcal{B}}\) if and only if the hyperplane \(x_i - x_j = a_{ij}\) is in \(\mathcal{B}\). Since the \(a_{ij}\)s are generic, the subarrangement \(\mathcal{B}\) is central if and only if the corresponding graph \(G_{\mathcal{B}}\) is a forest. For such a \(\mathcal{B}\), it is clear that \(|\mathcal{B}| = r(\mathcal{B}) = e(G_{\mathcal{B}})\). Hence,
\[
\overline{\chi}_{\mathcal{B}^n}(q, t) = \sum_{\mathcal{B} \subseteq \mathcal{B}^n \text{ central}} q^{r(\mathcal{B})}(t-1)^{|\mathcal{B}|} = \sum_{F \text{ forest on } [n]} q^{(n-1)-e(F)}(t-1)^{e(F)},
\]
proving the first claim. Now let \(c(F) = n - e(F)\) be the number of connected components of \(F\). We have
\[
1 + q \sum_{n \geq 1} \overline{\chi}_{\mathcal{B}^n}(q, t) \frac{x^n}{n!} = \sum_{n \geq 0} \sum_{F \text{ forest on } [n]} \left(\frac{q}{t-1}\right)^{c(F)} \frac{x^n(t-1)^n}{n!}
\]
\[
= \exp\left(\frac{q}{t-1} \sum_{n \geq 0} n^{n-2} \frac{x^n(t-1)^n}{n!}\right),
\]
using the exponential formula for exponential generating functions, and the fact that there are \(n^{n-2}\) trees on \(n\) labeled vertices [Stanley 1999]. \(\Box\)

**Theorem 4.5.** The threshold arrangement \(\mathcal{T}_n\) in \(\mathbb{R}^n\) consists of the hyperplanes \(x_i + x_j = 0\), for \(1 \leq i < j \leq n\). For all \(n \geq 0\) we have
\[
\overline{\chi}_{\mathcal{T}_n}(q, t) = \sum_{G \text{ graph on } [n]} q^{bc(G)}(t-1)^{e(G)},
\]

\(^1\)The \(a_{ij}\) are “generic” if no \(n\) of the hyperplanes have a nonempty intersection, and any nonempty intersection of \(k\) hyperplanes has rank \(k\). This can be achieved, for example, by requiring that the \(a_{ij}\)s are linearly independent over the rational numbers. Almost all choices of the \(a_{ij}\)s are generic.
where $bc(G)$ is the number of connected components of $G$ which are bipartite, and $e(G)$ is the number of edges of $G$. Also,

$$
\sum_{n \geq 0} \overline{X}_{\mathcal{B}}(q,t) \frac{x^n}{n!} = \left( \sum_{n \geq 0} \sum_{k=0}^{n} \binom{n}{k} k^{n-k} \frac{x^n}{n!} \right)^{q-1/2} \left( \sum_{n \geq 0} t^{\frac{n}{2}} \frac{x^n}{n!} \right).
$$

**Proof.** Once again, the proof of the first claim is easier using the definition of the coboundary polynomial. Every subarrangement $\mathcal{B}$ of $\mathcal{T}_n$ is central, and we can assign to it a graph $G_{\mathcal{B}}$ as in the proof of Theorem 4.4. In view of (3-2), we only need to check that $r(\mathcal{B}) = n - bc(G_{\mathcal{B}})$ and $|\mathcal{B}| = e(G_{\mathcal{B}})$. The second claim is trivial. To prove the first one, we show that $\dim(\bigcap \mathcal{B}) = bc(G_{\mathcal{B}})$.

Consider a point $p$ in $\bigcap \mathcal{B}$. We know that, if $ab$ is an edge in $G_{\mathcal{B}}$, then $p_a = -p_b$. If vertex $i$ is in a connected component $C$ of $G_{\mathcal{B}}$, then the value of $p_i$ determines the value of $p_j$ for all $j$ in $C$: $p_j = p_i$ if there is a path of even length between $i$ and $j$, and $p_j = -p_i$ if there is a path of odd length between $i$ and $j$. If $C$ is bipartite, this determines the values of the $p_j$s consistently. If $C$ is not bipartite, take a cycle of odd length and a vertex $k$ in it. We get that $p_k = -p_k$, so $p_k = 0$; therefore we must have $p_j = 0$ for all $j \in C$.

Therefore, to specify a point $p$ in $\bigcap \mathcal{B}$, we split $G_{\mathcal{B}}$ into its connected components. We know that $p_i = 0$ for all $i$ in connected components which are not bipartite. To determine the remaining coordinates of $p$ we have to specify the value of $p_j$ for exactly one $j$ in each bipartite connected component. Therefore $\dim(\bigcap \mathcal{B}) = bc(G_{\mathcal{B}})$, as desired.

From this point, it is possible to prove the second claim of Theorem 4.5 using the compositional formula for exponential generating functions, in the same way that we proved Theorem 4.4. However, the work involved is considerable, and it is much simpler to use our finite field method, Theorem 3.3, in this case. The proof that we obtain is very similar to the proofs of Theorems 4.1, 4.2 and 4.3, so we omit the details. □

5. Deformations of the braid arrangement

This section concerns the deformations of the braid arrangement of the form

$$
(5-1) \quad x_i - x_j = a_1, \ldots, a_k \quad 1 \leq i < j \leq n,
$$

where $A = \{a_1, \ldots, a_k\}$ is a fixed set of integers. Such arrangements have been studied extensively by Athanasiadis [2000] and Postnikov and Stanley [2000]. In this section we study the problem of finding their coboundary polynomials.

We proceed as follows. In the next subsection we introduce the family of graded $A$-graphs, and show in Proposition 5.6 that enumerating them is equivalent to the
problem at hand. By understanding the structure of those graphs, we obtain Theorem 5.7, a formula for the generating function of our coboundary polynomials.

The formula provided by Theorem 5.7 is not very explicit, as one might expect from the fact that it applies to such a large family of arrangements. However, we show that for $A \subseteq \{-1, 0, 1\}$, $A$-graphs possess additional structure, which makes it possible to obtain very explicit answers. This is done starting on page 19 for the Linial, Shi, semiorder, and Catalan arrangements. As a consequence, we also obtain new formulas for the generating functions of alternating trees, labelled trees, semiorders, and Dyck paths.

**Enumerating graphs to compute coboundary polynomials.**

**Definition 5.1.** An exponential sequence of arrangements $\mathcal{E} = (\mathcal{E}_0, \mathcal{E}_1, \ldots)$ is a sequence of arrangements satisfying the following properties:

1. $\mathcal{E}_n$ is an arrangement in $k^n$, for some fixed field $k$.
2. Every hyperplane in $\mathcal{E}_n$ is parallel to some hyperplane in the braid arrangement $A_n$.
3. For any subset $S$ of $[n]$, the subarrangement $\mathcal{E}_n^S \subseteq \mathcal{E}_n$, which consists of the hyperplanes in $\mathcal{E}_n$ of the form $x_i - x_j = c$ with $i, j \in S$, is isomorphic to the arrangement $\mathcal{E}_{|S|}$.

The special case $t = 0$ of the next result is due to Stanley [1996, Theorem 1.2]; we omit the proof, which is an easy extension of his.

**Theorem 5.2.** Let $\mathcal{E} = (\mathcal{E}_0, \mathcal{E}_1, \ldots)$ be an exponential sequence of arrangements. Then

$$1 + q \sum_{n \geq 1} \chi_{\mathcal{E}_n}(q, t) \frac{x^n}{n!} = \left( \sum_{n \geq 0} \chi_{\mathcal{E}_n}(1, t) \frac{x^n}{n!} \right)^q.$$

The most natural examples of exponential sequences of arrangements are the following. Fix a set $A$ of $k$ distinct integers $a_1 < \ldots < a_k$. Let $\mathcal{E}_n$ be the arrangement in $\mathbb{R}^n$ consisting of the hyperplanes

$$x_i - x_j = a_1, \ldots, a_k \quad 1 \leq i < j \leq n.$$

Then $(\mathcal{E}_0, \mathcal{E}_1, \ldots)$ is an exponential sequence of arrangements and Theorem 5.2 applies to this case. In fact, we can say much more about this type of arrangement. After proving the results in this section, we found out that Postnikov and Stanley [2000] had used similar techniques in computing the characteristic polynomials of these types of arrangements. Therefore, for consistency, we will use the terminology that they introduced.
Consider an arrangement of hyperplanes in \( E \). Let \( r \) be a nonnegative integer. Let the \( r \)th level of \( G \) be the set of vertices \( v \) such that \( h(v) = r \). Say \( G \) is planted if each one of its connected components has a vertex on the 0th level.

We will drop the subscripts when the underlying graded graph is clear. We will refer to \( h(v) \) as the height of vertex \( v \). The height of \( G \), denoted \( h(G) \), is the largest height of a vertex of \( G \).

**Definition 5.4.** Let \( G \) be a graded graph and \( r \) be a nonnegative integer. Let the \( r \)th level of \( G \) be the set of vertices \( v \) such that \( h(v) = r \). Say \( G \) is planted if each one of its connected components has a vertex on the 0th level.

**Definition 5.5.** If \( u < v \) are connected by edge \( e \) in a graded graph \( G \), let the type of \( e \) be \( s(e) = h(u) - h(v) \). Say \( G \) is an \( A \)-graph if the types of all edges of \( G \) are in \( A = \{a_1, \ldots, a_k\} \).

Recall that, for a graph \( G \), we let \( e(G) \) be the number of edges and \( v(G) \) be the number of connected components of \( G \).

**Proposition 5.6.** Let \( A = \{a_1, \ldots, a_k\} \), and let \( \mathcal{E}_n \) be the arrangement

\[
x_i - x_j = a_1, \ldots, a_k \quad 1 \leq i < j \leq n.
\]

Then, for \( n \geq 1 \),

\[
q \mathcal{Z}_{\mathcal{E}_n}(q, t) = \sum_G q^{e(G)}(t - 1)^{e(G)},
\]

where the sum is over all planted graded \( A \)-graphs on \([n]\).

**Proof.** We associate to each planted graded \( A \)-graph \( G = (V, E, h) \) on \([n]\) a central subarrangement \( \mathcal{A}_G \) of \( \mathcal{E}_n \). It consists of the hyperplanes \( x_i - x_j = h(i) - h(j) \), for each \( i < j \) such that \( ij \) is an edge in \( G \). This is a subarrangement of \( \mathcal{E}_n \) because \( h(i) - h(j) \), the type of edge \( ij \), is in \( A \). It is central because the point \((h(1), \ldots, h(n)) \in \mathbb{R}^n \) belongs to all these hyperplanes.

**Example.** Consider an arrangement \( \mathcal{E}_8 \) in \( \mathbb{R}^8 \), with a subarrangement consisting of the hyperplanes \( x_1 - x_2 = 4, x_1 - x_3 = -1, x_1 - x_6 = 0, x_1 - x_8 = 1, x_2 - x_3 = -5 \) and \( x_4 - x_7 = 2 \). Figure 1 shows the planted graded \( A \)-graph corresponding to this subarrangement.

This is in fact a bijection between planted graded \( A \)-graphs on \([n]\) and central subarrangements of \( \mathcal{E}_n \). To see this, take a central subarrangement \( \mathcal{A} \). We will recover the planted graded \( A \)-graph \( G \) that it came from. For each pair \((i, j)\) with \( 1 \leq i < j \leq n \), \( \mathcal{A} \) can have at most one hyperplane of the form \( x_i - x_j = a_i \). If this hyperplane is in \( \mathcal{A} \), we must put edge \( ij \) in \( G \), and demand that the heights \( h(i) \) and \( h(j) \) satisfy \( h(i) - h(j) = a_i \). When we do this for all the hyperplanes in \( \mathcal{A} \), the height requirements that we introduce are consistent, because \( \mathcal{A} \) is central.
However, these requirements do not fully determine the heights of the vertices; they only determine the relative heights within each connected component of $G$. Since we want $G$ to be planted, we demand that the vertices with the lowest height in each connected component of $G$ should have height 0. This does determine $G$ completely, and clearly $\mathcal{A} = \mathcal{A}_G$.

With this bijection in hand, and keeping (3-2) in mind, it remains to show that $r(\mathcal{A}_G) = n - c(G)$ and $|\mathcal{A}_G| = e(G)$. The second of these claims is trivial. We omit the proof of the first one which is very similar to, and simpler than, that of $r(\mathcal{B}) = n - b c(G_{\mathcal{B}})$ in our proof of Theorem 4.5. \hfill \blacksquare

**Theorem 5.7.** Let $A = \{a_1, \ldots, a_n\}$ and let $\mathcal{E}_n$ be the arrangement
\[ x_i - x_j = a_1, \ldots, a_k \quad 1 \leq i < j \leq n. \]

Let
\[ A_r(t, x) = \sum_{n \geq 0} \left( \sum_{f : [n] \to [r]} a^{(f)} \right) \frac{x^n}{n!}, \]
where the inside sum is over all functions $f : [n] \to [r]$, and
\[ a(f) = \#\{(i, j) \mid 1 \leq i < j \leq n, \ f(i) - f(j) \in A\}. \]

Then
\[ 1 + q \sum_{n \geq 1} \mathcal{K}_{\mathcal{E}_n}(q, t) \frac{x^n}{n!} = \left( \lim_{r \to \infty} \frac{A_r(t, x)}{A_{r-1}(t, x)} \right)^q. \]

**Remark.** The limit in (5-4) is a limit in the sense of convergence of formal power series. Let $F_1(t, x), F_2(t, x), \ldots$ be a sequence of formal power series. We say that $\lim_{n \to \infty} F_n(t, x) = F(t, x)$ if for all $a$ and $b$ there exists a constant $N(a, b)$ such that, for all $n$ larger than $N(a, b)$, the coefficient of $t^a x^b$ in $F_n(t, x)$ is equal
to the coefficient of $t^a x^b$ in $F(t, x)$. For more on this notion of convergence, see [Niven 1969] or [Stanley 1997, Section 1.1].

**Proof of Theorem 5.7.** Let $v(G)$ be the number of vertices of graph $G$. First we prove that

\[(5-5) \quad A_r(t, x) = \sum_G (t - 1)^{e(G)} \frac{x^{v(G)}}{v(G)!} \]

where the sum is over all graded $A$-graphs $G$ on $[n]$ of height less than $r$. The coefficient of $\frac{x^n}{n!}$ in the right-hand side of (5-5) is $\sum_G (t - 1)^{e(G)}$, summing over all graded $A$-graphs $G$ on $[n]$ with height less than $r$. We have

\[
\sum_G (t - 1)^{e(G)} = \sum_{h:[n] \to [0, r-1]} \sum_{G: h_G = h} (t - 1)^{e(G)} = \sum_{h:[n] \to [0, r-1]} (1 + (t - 1))^{a(h)} = \sum_{f:[n] \to [r]} t^{a(f)}.
\]

The only tricky step here is the second: if we want all graded $A$-graphs $G$ on $[n]$ with a specified grading $h$, we need to consider the possible choices of edges of the graph. Any edge $ij$ can belong to the graph, as long as $h(i) - h(j) \in A$, so there are $a(h)$ possible edges.

Equation (5-5) suggests the following definitions. Let

\[
B_r(t, x) = \sum_G t^{e(G)} \frac{x^{v(G)}}{v(G)!}
\]

where the sum is over all *planted* graded $A$-graphs $G$ of height less than $r$, and let

\[
B(t, x) = \sum_G t^{e(G)} \frac{x^{v(G)}}{v(G)!}
\]

where the sum is over all *planted* graded $A$-graphs $G$.

The equation

\[(5-6) \quad 1 + q \sum_{n \geq 1} \frac{\chi(a_n(q, t))}{n!} \frac{x^n}{n!} = B(t - 1, x)^q,
\]

follows from Proposition 5.6, using either Theorem 5.2 or the compositional formula for exponential generating functions.

Now we claim that $B(t, x) = \lim_{r \to \infty} B_r(t, x)$. In a planted graded $A$-graph $G$ with $e$ edges and $v$ vertices, each vertex has a path of length at most $v$ that connects it to a vertex on the 0th level. Therefore $h(G) \leq v \cdot \max(|a_1|, \ldots, |a_k|)$, so the coefficients of $t^e x^v$ in $B_r(t, x)$ and $B(t, x)$ are equal for $r > v \cdot \max(|a_1|, \ldots, |a_k|)$. 


Then it is not difficult to show that

\[(5-7) \quad B(t - 1, x) = \lim_{r \to \infty} B_r(t - 1, x).\]

Here it is necessary to check that \(B(t - 1, x)\) is, indeed, a formal power series. This follows from the observation that the coefficient of \(x^n\) in \(B(t, x)\) is a polynomial in \(t\) of degree at most \(\binom{n}{2}\). Once again, see [Stanley 1997, Section 1.1] for more information on these technical details.

Next, we show that

\[(5-8) \quad B_r(t - 1, x) = A_r(t, x)/A_{r-1}(t, x)\]

or, equivalently, that \(A_r(t, x) = B_r(t - 1, x)A_{r-1}(t, x)\). The multiplication formula for exponential generating functions [Stanley 1999, Proposition 5.1.1] and (5-5) give us a combinatorial interpretation of this identity. We need to show that the ways of putting the structure of a graded \(A\)-graph \(G\) with \(h(G) < r\) on \([n]\) can be put in correspondence with the ways of doing the following: first splitting \([n]\) into two disjoint sets \(S_1\) and \(S_2\), then putting the structure of a *planted* graded \(A\)-graph \(G_1\) with \(h(G_1) < r\) on \(S_1\), and then putting the structure of a graded \(A\)-graph \(G_2\) with \(h(G_2) < r - 1\) on \(S_2\). We also need that, in that correspondence, \((t - 1)^{e(G)} = (t - 1)^{e(G_1)}(t - 1)^{e(G_2)}\).

We do this as follows. Let \(G\) be a graded \(A\)-graph \(G\) with \(h(G) < r\). Let \(G_1\) be the union of the connected components of \(G\) which contain a vertex on the 0th level. Put a grading on \(G_1\) by defining \(h_{G_1}(v) = h_G(v)\) for \(v \in G_1\). Let \(G_2 = G - G_1\). It is clear that \(h_{G_2}(v) \geq 1\) for all \(v \in G_2\); therefore we can put a grading on \(G_2\) by defining \(h_{G_2}(v) = h_G(v) - 1\) for \(v \in G_2\). Then \(G_1\) is a planted

\[\text{Figure 2. The decomposition of a graded } A\text{-graph.}\]
graded A-graph with \( h(G_1) < r \), \( G_2 \) is a graded A-graph with \( h(G_2) < r - 1 \), and \((t - 1)^{e(G)} = (t - 1)^{e(G_1)}(t - 1)^{e(G_2)}\). It is clear how to recover \( G \) from \( G_1 \) and \( G_2 \). Figure 2 illustrates this bijection with an example.

Now we just have to put together (5-6), (5-7) and (5-8) to complete the proof of Theorem 5.7. \( \square \)

**Subarrangements of the Catalan arrangement.** The Catalan arrangement \( C_n \) in \( \mathbb{R}^n \) consists of the hyperplanes

\[
x_i - x_j = -1, 0, 1 \quad 1 \leq i < j \leq n.
\]

When the arrangement in Theorem 5.7 is a subarrangement of the Catalan arrangement, we can say more about the power series \( A_r \) of (5-3). Let

\[
A(t, x, y) = \sum_r A_r(t, x)y^r = \sum_{n \geq 0} \left( \sum_{f : [n] \to [r]} t^{a(f)} \right) \frac{x^n}{n!} y^r
\]

and let

\[
S(t, x, y) = \sum_{n \geq 0} \left( \sum_{f : [n] \to [r]} t^{a(f)} \right) \frac{x^n}{n!} y^r
\]

where the inner sum is over all surjective functions \( f : [n] \to [r] \). The following proposition reduces the computation of \( A(t, x, y) \) to the computation of \( S(t, x, y) \), which is often easier in practice.

**Proposition 5.8.** If \( A \subseteq \{-1, 0, 1\} \) in the notation of Theorem 5.7, we have

\[
A(t, x, y) = \frac{S(t, x, y)}{1 - y S(t, x, y)}
\]

**Proof.** Once again, we think of this as an identity about exponential generating functions in the variable \( x \). Fix \( n, r \), and a function \( f : [n] \to [r] \). Write \([r] - \text{Im } f = \{m_1, m_1 + m_2, \ldots, m_1 + \cdots + m_{k-1}\} \), so the image of \( f \) is \( M_1 \cup \cdots \cup M_k = \{1, \ldots, m_1 - 1\} \cup \{m_1 + 1, \ldots, m_1 + m_2 - 1\} \cup \cdots \cup \{m_1 + \cdots + m_{k-1} + 1, \ldots, m_1 + \cdots + m_k - 1\} \). Here \( m_1, \ldots, m_k \) are arbitrary positive integers such that \( m_1 + \cdots + m_k - 1 = r \). For \( 1 \leq i \leq k \), let \( f_i \) be the restriction of \( f \) to \( f^{-1}(M_i) \); it maps \( f^{-1}(M_i) \) surjectively to \( M_i \). Then we can “decompose” \( f \) in a unique way into the \( k \) surjective functions \( f_1, \ldots, f_k \). The weight \( w(f) \) corresponding to \( f \) in \( A(t, x, y) \) is \( t^{a(f)} y^r \), while the weight \( w(f_i) \) corresponding to \( f_i \) in \( S(t, x, y) \) is \( t^{a(f_i)} y^{|m_i| - 1} \).

Now observe that \( a(f) = a(f_1) + \cdots + a(f_k) \): whenever we have a pair of numbers \( 1 \leq i < j \leq n \) counted by \( a(f) \), since \( f(i) - f(j) \in \{-1, 0, 1\} \), we know that \( f(i) \) and \( f(j) \) must be in the same \( M_h \). Therefore \( i \) and \( j \) are in the same \( f^{-1}(M_h) \), and this pair is also counted in \( a(f_h) \). We also have that \( r = \cdots \)
\[(m_1 - 1) + \cdots + (m_k - 1) + (k - 1). \] Therefore \( w(f) = w(f_1) \cdots w(f_k) y^{k-1}. \) It follows from the compositional formula for exponential generating functions that

\[
A(t, x, y) = \sum_{k \geq 1} S(t, x, y) y^{k-1} = \frac{S(t, x, y)}{1 - y S(t, x, y)},
\]
as desired.

Considering the different subsets of \([-1, 0, 1]\), we get six nonisomorphic subarrangements of the Catalan arrangement. They come from the subsets \(\emptyset, \{0\}, \{1\}, \{-1, 1\}\) and \([-1, 0, 1]\). The corresponding subarrangements are the empty arrangement, the braid arrangement, the Linial arrangement, the Shi arrangement, the semiorder arrangement, and the Catalan arrangement, respectively. The empty arrangement is trivial, and the braid arrangement was already treated in detail starting on page 10. We now have a technique that lets us talk about the remaining four arrangements under the same framework. We will do this in the remainder of this section.

**The Linial arrangement.** The Linial arrangement \(\mathcal{L}_n\) consists of the hyperplanes \(x_i - x_j = 1\) for \(1 \leq i < j \leq n\). This arrangement was first considered by Linial and Ravid. It was later studied by Athanasiadis [1996] and Postnikov and Stanley [2000], who independently computed the characteristic polynomial of \(\mathcal{L}_n\):

\[
\chi_{\mathcal{L}_n}(q) = \frac{q}{2^n} \sum_{k=0}^{n} \binom{n}{k} (q-k)^{n-1}.
\]

They also put the regions of \(\mathcal{L}_n\) in bijection with several different sets of combinatorial objects. Perhaps the simplest such set is the set of **alternating trees** on \([n+1]\): the trees such that every vertex is either larger or smaller than all its neighbors.

Now we present the consequences of Proposition 5.8 and Propositions 5.6 and 5.7 for the Linial arrangement. Say that a poset \(P\) on \([n]\) is **naturally labeled** if \(i < j\) in \(P\) implies \(i < j\) in \(\mathbb{Z}^+\).

**Proposition 5.9.** For all \(n \geq 1\) we have

\[
q \chi_{\mathcal{L}_n}(q, t) = \sum_P q^{c(P)} (t-1)^{e(P)}
\]

where the sum is over all naturally labeled, graded posets \(P\) on \([n]\). Here \(c(P)\) and \(e(P)\) denote the number of components and edges of the Hasse diagram of \(P\), respectively.

**Proof.** There is an obvious bijection between Hasse diagrams of naturally labeled graded posets on \([n]\) and planted graded \([1]\)-graphs on \([n]\). The result then follows immediately from Proposition 5.6. \(\square\)
Theorem 5.10. Let
\[ A_r(t, x) = \sum_{n \geq 0} \left( \sum_{f : [n] \rightarrow [r]} i^{id(f)} \right) \frac{x^n}{n!}, \]
where \( id(f) \) denotes the number of inverse descents of the word \( f(1) \ldots f(n) \), that is, the number of pairs \( i, j \) with \( 1 \leq i < j \leq n \) such that \( f(i) - f(j) = 1 \). Then
\[ 1 + q \sum_{n \geq 1} \chi_{\mathcal{L}_n}(q, t) \frac{x^n}{n!} = \left( \lim_{r \to \infty} \frac{A_r(t, x)}{A_{r-1}(t, x)} \right)^q. \]

Proof: This is immediate from Theorem 5.7. \( \square \)

Recall that the descents of a permutation \( \sigma = \sigma_1 \ldots \sigma_r \in S_r \) are the indices \( i \) such that \( \sigma_i > \sigma_{i+1} \). For more information about descents, see [Stanley 1997, Section 1.3], for example. We call \( id(f) \) the number of inverse descents, because they generalize descents in the following way. If \( \pi : [r] \rightarrow [r] \) is a permutation, then \( id(\pi) \) is the number of descents of the permutation \( \pi^{-1} \).

We can use Theorem 5.10 to say more about the characteristic polynomial of \( \mathcal{L}_n \) which, as discussed on page 4, is given by \( \chi_{\mathcal{L}_n}(q) = q \chi_{\mathcal{L}_n}(q, 0) \).

Theorem 5.11. Let
\[ 1 + q \sum_{n \geq 1} \chi_{\mathcal{L}_n}(q, t) \frac{x^n}{n!} = \left( \lim_{r \to \infty} \frac{A_r(t, x)}{A_{r-1}(t, x)} \right)^q. \]

Then we have
\[ \sum_{n \geq 0} \chi_{\mathcal{L}_n}(q) \frac{x^n}{n!} = \left( \lim_{r \to \infty} \frac{A_r(x)}{A_{r-1}(x)} \right)^q. \]

In particular, if \( f_n \) is the number of alternating trees on \([n + 1]\), we have
\[ \sum_{n \geq 0} (-1)^n f_n \frac{x^n}{n!} = \lim_{r \to \infty} \frac{A_{r-1}(x)}{A_r(x)}. \]

Proof: In view of Theorem 5.10 and Proposition 5.8, we compute \( S(0, x, y) \). From (5-9), the coefficient of \( \frac{x^n}{n!} y^r \) in \( S(0, x, y) \) is equal to the number of surjective functions \( f : [n] \rightarrow [r] \) with no inverse descents. These are just the nondecreasing surjective functions \( f : [n] \rightarrow [r] \). For \( n \geq 1 \) and \( r \geq 1 \) there are \( \binom{n-1}{r-1} \) such functions, and for \( n = r = 0 \) there is one such function. In the other cases there are none. Therefore
\[ S(0, x, y) = 1 + \sum_{n \geq 1} \sum_{r \geq 1} \binom{n-1}{r-1} \frac{x^n}{n!} y^r = 1 + \sum_{n \geq 1} \frac{x^n}{n!} y(1 + y)^{n-1} = \frac{1 + ye^{(1+y)}}{1 + y}. \]
Proposition 5.8 then implies that
\[ A(0, x, y) = \frac{1 + ye^{x(1+y)}}{1 - y^2 e^{x(1+y)}}, \]
in agreement with (5-10), and the theorem follows. \(\square\)

The Shi arrangement. The Shi arrangement \(Y_n\) consists of the hyperplanes \(x_i - x_j = 0\), 1 for \(1 \leq i < j \leq n\). Shi [1986, Chapter 7; 1987] first considered this arrangement, and showed that it has \((n + 1)^{n-1}\) regions. Headley [1994, Chapter VI; 1997] later computed the characteristic polynomial of \(Y_n\):
\[ \chi_{Y_n}(q) = q(q - n)^{n-1}. \]
Stanley [1996; 1998b] gave a nice bijection between regions of the Shi arrangement and parking functions of length \(n\). Parking functions were first introduced by Konheim and Weiss [1966]; for more information, see [Stanley 1999, Exercise 5.49].

For the Shi arrangement, we can say the following.

**Theorem 5.12.** Let
\[ A_r(x) = \sum_{n=0}^{r} (r - n)^n \frac{x^n}{n!}. \]
Then we have
\[ \sum_{n \geq 0} \chi_{Y_n}(q) \frac{x^n}{n!} = \left( \lim_{r \to \infty} \frac{A_r(x)}{A_{r-1}(x)} \right)^q. \]
In particular, we have
\[ \sum_{n \geq 0} (-1)^n (n + 1)^{n-1} \frac{x^n}{n!} = \lim_{r \to \infty} \frac{A_{r-1}(x)}{A_r(x)}. \]

**Proof:** We proceed in the same way that we did in Theorem 5.11. In this case, we need to compute the number of surjective functions \(f : [n] \to [r]\) such that \(f(i) - f(j)\) is never equal to 0 or 1 for \(i < j\). These are just the surjective, strictly increasing functions. There is only one of them when \(n = r\), and there are none when \(n \neq r\). Hence
\[ S(0, x, y) = \sum_{n \geq 0} \frac{x^n}{n!} y^n = e^{xy}. \]
The rest follows easily by computing \(A(0, x, y)\) and \(A_r(x)\) explicitly. \(\square\)
The semiorder arrangement. The semiorder arrangement $\mathcal{S}_n$ consists of the hyperplanes $x_i - x_j = -1$, for $1 \leq i < j \leq n$. A semiorder on $[n]$ is a poset $P$ on $[n]$ for which there exist $n$ unit intervals $I_1, \ldots, I_n$ of $\mathbb{R}$, such that $i < j$ in $P$ if and only if $I_i$ is disjoint from $I_j$ and to the left of it. It is known [Scott and Suppes 1958] that a poset is a semiorder if and only if it does not contain a subposet isomorphic to $3 + 1$ or $2 + 2$. We are interested in semiorders because the number of regions of $\mathcal{S}_n$ is equal to the number of semiorders on $[n]$, as shown in [Postnikov and Stanley 2000] and [Stanley 1996].

**Theorem 5.13.** Let

$$\frac{1 - y + ye^x}{1 - y + y^2 - ye^x} = \sum_{r \geq 0} A_r(x)y^r.$$

Then we have

$$\sum_{n \geq 0} \chi_{\mathcal{S}_n}(q) \frac{x^n}{n!} = \left( \lim_{r \to \infty} \frac{A_r(x)}{A_{r-1}(x)} \right)^q.$$

In particular, if $i_n$ is the number of semiorders on $[n]$, we have

$$\sum_{n \geq 0} (-1)^ni_n \frac{x^n}{n!} = \lim_{r \to \infty} \frac{A_{r-1}(x)}{A_r(x)}.$$

**Proof.** In this case, $S(0, x, y)$ counts surjective functions $f : [n] \to [r]$ such that $f(i) - f(j)$ is never equal to 1 for $i \neq j$. Such a function has to be constant; so it can only exist (and is unique) if $n \geq 1$ and $r = 1$ or if $n = r = 0$. Thus

$$S(0, x, y) = 1 + (e^x - 1)y$$

and the rest follows easily. \qed

The Catalan arrangement. The Catalan arrangement $C_n$ consists of the hyperplanes $x_i - x_j = -1, 0, 1$ for $1 \leq i < j \leq n$. Stanley [1996] observed that the number of regions of this arrangement is $n!C_n$, where $C_n = \frac{1}{n+1} \binom{2n}{n}$ is the $n$-th Catalan number. For (much) more information on the Catalan numbers; see [Stanley 1999, Chapter 6], especially Exercise 6.19.

**Theorem 5.14.** Let

$$A_r(x) = \sum_{n=0}^{\left\lfloor \frac{r+1}{2} \right\rfloor} \binom{r - n + 1}{n} x^n.$$

Then we have

$$\sum_{n \geq 0} \chi_{C_n}(q) \frac{x^n}{n!} = \left( \lim_{r \to \infty} \frac{A_r(x)}{A_{r-1}(x)} \right)^q.$$
In particular,

\[
\frac{\sqrt{1 + 4x} - 1}{2x} = \sum_{n \geq 0} (-1)^n C_n x^n = \lim_{r \to \infty} \frac{A_{r-1}(x)}{A_r(x)}.
\]

**Proof.** If \( f : [n] \to [r] \) is a surjective function such that \( f(i) - f(j) \) is never equal to \(-1, 0, \) or \( 1 \) for \( i \neq j \), then \( n = r = 0 \) or \( n = r = 1 \). Thus \( S(x, y, 0) = 1 + xy \), and the rest of the proof is straightforward. \( \square \)

The polynomial \( A_r(x) \) is a simple transformation of the Fibonacci polynomial. The number of words of length \( r \), consisting of 0s and 1s, which do not contain two consecutive 1s, is equal to \( F_{r+2} \), the \((r+2)\)th Fibonacci number. It is easy to see that the polynomial \( A_r(x) \) counts those words according to the number of 1s they contain. In particular, \( A_r(1) = F_{r+2} \).

We close with an amusing observation. Irresponsibly\(^2\) plugging \( x = 1 \) into (5-11), we obtain an unconventional “proof” of the asymptotic rate of growth of Fibonacci numbers:

\[
\frac{\sqrt{5} - 1}{2} = \lim_{r \to \infty} \frac{F_{r-1}}{F_r}.
\]

6. Acknowledgments

This work is Chapter 2 of my Ph.D. thesis [Ardila 2003]. I would like to thank my advisor, Richard Stanley, for introducing me to the topic of hyperplane arrangements, and for asking some of the questions which led to this investigation. I am also grateful to Ira Gessel and Vic Reiner for helpful discussions on this subject. Finally, I thank the referee for a careful reading of the manuscript, and valuable suggestions for improvement.

**References**


\(^2\)We are not necessarily justified in doing this, since we have only proved equality in (5-11) as formal power series!
Federico Ardila
Mathematics Department
San Francisco State University
1600 Holloway Avenue
San Francisco, CA 94110
United States
federico@math.sfsu.edu
http://math.sfsu.edu/federico


Received September 10, 2004.
ON THE UNIT GROUP OF SOME MULTIQUADRATIC NUMBER FIELDS

ELLIO T BENJAM IN, FRANZ L EM M E R MEYER AND CHIP SNYDER

We study the index of the group of units in the genus field of an imaginary quadratic number field modulo the subgroup generated by the units of the quadratic subfields (over \( \mathbb{Q} \)) of the genus field.

1. Introduction

One major problem in algebraic number theory is the computation of the class number \( h(K) \) for a number field \( K \). In the case of quadratic fields, this problem is easily solved by elementary methods. Once the field degree is larger than 2, the problem becomes more challenging. Historically, the oldest case after the quadratic fields seems to be when \( K \) runs through a particular family of quartic bicyclic fields over \( \mathbb{Q} \), meaning that \( \text{Gal}(K/\mathbb{Q}) \cong (2, 2) \) (here \( (a_1, \ldots, a_r) \) denotes the direct sum of cyclic groups of order \( a_i \), for \( i = 1, \ldots, r \)). Dirichlet [1842] in essence computed the class number \( h(K) \) for the family of quartic fields \( K = \mathbb{Q}(\sqrt{-1}, \sqrt{m}) \), \( m \) a positive nonsquare integer. Namely, let \( k_1 = \mathbb{Q}(\sqrt{-1}) \), \( k_2 = \mathbb{Q}(\sqrt{m}) \), and \( k_3 = \mathbb{Q}(\sqrt{-m}) \), and denote by \( E_F \) the group of units of a number field \( F \). Then Dirichlet discovered the class number formula

\[
h(K) = \frac{1}{2} q(K/\mathbb{Q}) h(k_2) h(k_3),
\]

where \( q = q(K/\mathbb{Q}) = (E_K : E_{k_1} E_{k_2} E_{k_3}) \). Dirichlet went on to show that the unit index \( q \) could be determined and was equal to 1 or 2.

Over time, Dirichlet’s formula has been generalized in several directions; see in particular [Herglotz 1922; Kubota 1953; 1956; Kuroda 1950; Lemmermeyer 1994b; Wada 1966], and references therein. One particularly striking formula is usually attributed to Kuroda [1950], but in fact goes back to Herglotz [1922] in an equivalent, if less convenient, form for \( q \). Let \( L = \prod_i k_i \) be the multiquadratic field generated as the composite of all its quadratic subfields \( k_i \), and suppose further that \( [L : \mathbb{Q}] = 2^m \). Then

\[
h(L) = \frac{1}{2^m} q(L/\mathbb{Q}) \prod_i h(k_i),
\]
where \( q = q(L/\mathbb{Q}) = (E_L : \prod_i E_{k_i}) \) and

\[
v = \begin{cases} 
  m(2^m - 1) & \text{if } L \text{ is real,} \\
  (m - 1)(2^m - 2 - 1) + 2^m - 1 & \text{if } L \text{ is complex.}
\end{cases}
\]

Hence \( h(L) \) can be computed easily provided that the unit index \( q(L/\mathbb{Q}) \) can be computed. Herein lies the obstruction to an easy determination of the class number of multiquadratic number fields. For quartic bicyclic fields, Kubota [1956] gave a method for finding a system of fundamental units and thus for computing \( q \). Wada [1966], generalized Kubota’s method, creating an algorithm for computing fundamental units in any given multiquadratic field. However, in general there seems to be no explicit formula for \( q \), even when \( L \) is of degree 4 over \( \mathbb{Q} \).

This brings us to the purpose of this article. We try to glean some understanding of the difficulties in computing the unit index by giving explicit computations of \( q \) for special families of multiquadratic fields \( L \). We consider the special case of the genus field \( L = k_{\text{gen}} \) of a complex quadratic field \( k \) for which the 2-rank of the class group \( \text{Cl}(k) \) of \( k \) is \( \leq 3 \). (Recall that the \( 2^n \)-rank of a finite abelian group \( G \) is the minimal number of generators of the factor group \( G^{2^{n-1}}/G^{2^n} \).) If the 2-rank of \( \text{Cl}(k) \) is 1, then \( [L : \mathbb{Q}] = 4 \), by genus theory, and in this case it is known that \( q = 1 \) (see [Lemmermeyer 1995], for instance; the proof is easy — see the next section).

Next, if the 2-rank is 2, then \( [L : \mathbb{Q}] = 8 \) by genus theory. In this case, we reduce the problem to that of computing \( q(K/\mathbb{Q}) \) where \( K \) is the maximal real subfield of \( L \). But then \( K \) is a totally real bicyclic field and we may apply the results of [Kubota 1956] to compute \( q(K/\mathbb{Q}) \). We find that \( q(L/\mathbb{Q}) = 8 \) or 2 according as the 2-class field tower of \( k \) is of length 1 or \( > 1 \). (Here \( k^1 \) is the Hilbert 2-class field of \( k \) and \( k^{n+1} = (k^n)^1 \); the length of the 2-class field tower of \( k \) is the cardinality of the set of \( k^n \ ).)

For the case where the 2-rank of \( \text{Cl}(k) \) is 3, we seem to be in new territory. We restrict to the case of elementary 2-class group. Specifically, we assume \( \text{Cl}_2(k) \cong (2, 2, 2) \), so \( L = k_{\text{gen}} = k^1 \). If the rank of \( \text{Cl}_2(k^1) \) is 2 as a module over the integral group ring \( \Lambda = \mathbb{Z}[\text{Gal}(k^1/k)] \), then \( q(L/\mathbb{Q}) = 2^7 \). This condition on the \( \Lambda \)-rank is, by the way, a natural one; see [Benjamin et al. 2003]. We then obtain less complete information about \( q \) for the other case where \( \text{Cl}_2(k^1) \) is of \( \Lambda \)-rank 3. In the particular fields we consider, \( q = 2^4 \) or \( 2^5 \).

2. The Main Results

Let \( k \) be an imaginary quadratic field for which the 2-rank of \( \text{Cl}(k) \) is \( t - 1 \). Hence, by genus theory, \( k = \mathbb{Q}(\sqrt{d_1 \cdots d_t}) \), where disc \( k = d_1 \cdots d_t \) is a factorization of the discriminant of \( k \) into distinct prime discriminants \( d_i \) divisible by the rational prime \( p_i \) for \( i = 1, \ldots, t \). Then \( L = k_{\text{gen}} = \mathbb{Q}(\sqrt{d_1}, \ldots, \sqrt{d_t}) \) and hence multiquadratic
of degree $2^t$ over $\mathbb{Q}$. Hence for $t \geq 2$ Kuroda’s class number formula above yields

$$h(L) = \frac{1}{2^v} q(L/\mathbb{Q}) \prod_i h(k_i),$$

where $k_i$ range over the $2^t - 1$ quadratic subfields of $L$, $q(L/\mathbb{Q}) = (E_L : \prod_i E_{k_i})$, and $v = (t - 1)\left(2^{i-2} - 1\right) + 2^{i-1} - 1$ since $L$ is complex.

We start our computations of $q = q(L/\mathbb{Q})$ by first considering $t = 2$ (for the sake of completeness).

**Theorem 1.** Let $k$ be a complex quadratic field and $L = k_{\text{gen}}$. If the 2-rank of $\text{Cl}(k)$ equals 1, then $q(L/\mathbb{Q}) = 1$.

**Proof.** Since the 2-rank of $\text{Cl}(k)$ is 1, $t = 2$ so $k = \mathbb{Q}(\sqrt{d_1d_2})$ for prime discriminants $d_1, d_2$. Then $L = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2})$. Now, by Kuroda’s class number formula (where $t = 2$ implies $v = 1$),

$$h(L) = \frac{1}{2} q(L/\mathbb{Q}) h(d_1d_2) h(d_1) h(d_2),$$

where $h(n) = h(\mathbb{Q}(\sqrt{n}))$. Now, it is well known that $h(d)$ is odd for any prime discriminant $d$. Moreover, by the Artin map, $\text{Gal}(k^1/k) \simeq \text{Cl}_2(k)$ where $\text{Cl}_2(k)$ is the 2-class group of $k$, the Sylow 2-subgroup of the class group, the order of which is $h_2(k)$, the 2-class number of $k$. Now consider $G = \text{Gal}(k^2/k)$. Since the commutator subgroup $G' = \text{Gal}(k^2/k^1)$, we see $G/G' \simeq \text{Gal}(k^1/k) \simeq \text{Cl}_2(k)$. But in the present case, $\text{Cl}_2(k)$ is cyclic, whence $G' = \{1\}$, and thus $k^2 = k^1$. But then since in general $k^1 \subseteq L^1 \subseteq k^2$, we have $L^1 = k^1$. Therefore, $h_2(L) = [L^1 : L] = [k^1 : k] = [k^1 : L]/2 = h_2(k)/2$. Now by restricting to 2-class numbers and using the fact that $q$ is a power of two, (see [Wada 1966], for instance) the Kuroda class number formula becomes

$$h_2(L) = \frac{1}{2} q(L/\mathbb{Q}) h_2(k) h_2(d_1) h_2(d_2).$$

From the preceding discussion we get $\frac{1}{2} h_2(k) = \frac{1}{2} q h_2(k)$, as needed. \qed

Next, we consider the case where the 2-rank of $\text{Cl}(k)$ is 2, i.e. $t = 3$. Hence $k = \mathbb{Q}(\sqrt{d_1d_2d_3})$, with prime discriminants $d_i$. Moreover, since $k$ is complex, disc $k < 0$ so either all the $d_i$ are negative or exactly two are positive, say $d_1, d_2 > 0$, $d_3 < 0$. Notice that we have $L = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}, \sqrt{d_3})$. Let $K = L^+$ be the maximal real subfield of $L$, (so $K = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2})$ if say $d_1, d_2 > 0$, $d_3 < 0$, and $K = \mathbb{Q}(\sqrt{d_1d_2}, \sqrt{d_2d_3})$ if $d_i < 0$, for $i = 1, 2, 3$). But then it follows that $q(L/\mathbb{Q}) = Q(L/K) q(K/\mathbb{Q})$, where $Q(L/K) = (E_L : W_L E_K)$ with $W_L$ the group of roots of unity of $L$. To see this apply for example [Benjamin et al. 2003, Proposition 1], where we notice that any primitive eighth root $\zeta_8$ of unity is not contained in $L$ since any ramification index of a prime in $L/\mathbb{Q}$ must divide 2, whereas 2 is totally ramified in $\mathbb{Q}(\zeta_8)$. Now suppose $d_1, d_2 > 0$, $d_3 < 0$. By [Lemmermeyer
1995, Theorem 1], $Q(L/K) = 1$, since $L = K(\sqrt{d_3})$ implies that $L/K$ is essentially ramified if $d_3 \neq -4$ and that $2 \mathcal{O}_K$ is not an ideal square if $d_3 = -4$, see [Lemmermeyer 1995] again for the details. Thus we have $q(L/\mathbb{Q}) = q(K/\mathbb{Q})$. If however all the $d_i < 0$, then for any $i$, $L = K(\sqrt{d_i})$. In this case, it can be shown that $Q(L/K) = 2$ by [Lemmermeyer 1995], but we shall see that this is the case by another method. In either case, it is well known that $Q(L/K) = 1, 2$ (see [Hasse 1985, Satz 14]), and moreover, by Kubota [Kubota 1956], $q(K/\mathbb{Q})$ divides 4. Thus $q(L/\mathbb{Q})$ must divide 8.

**Theorem 2.** Let $k$ be a complex quadratic field with $2$-rank $\text{Cl}(k) = 2$. Then for $L = k_{\text{gen}}, q(L/\mathbb{Q}) = 8$ or 2 according as the 2-class field tower of $k$ is of length 1 or $> 1$.

**Proof.** By assumption, $k = \mathbb{Q}(\sqrt{d_1d_2d_3})$ for prime discriminants $d_i$. Now notice that $K_i = k(\sqrt{d_i})$ for $i = 1, 2, 3$ are the three unramified quadratic extensions of $k$ in $L$. These fields are quartic bicyclic extensions of $\mathbb{Q}$ and so Kuroda’s class number yields

$$h(K_1) = \frac{1}{2} q(K_1/\mathbb{Q})h(k)h(d_2d_3)h(d_1),$$

since $m = 2$, so $v = 1$, (analogously for $K_2$ and $K_3$). Now since the $K_i$ are unramified quadratic extensions of a complex quadratic field $k$, it is known that $q(K_i/\mathbb{Q}) = 1$; see for example [Lemmermeyer 1995]. Hence by considering 2-class numbers so that we may use $h_2(d_i) = 1$, we have

$$h_2(L) = \frac{1}{32} q(L/\mathbb{Q})h_2(k)h_2(d_1d_2)h_2(d_1d_3)h_2(d_2d_3), \quad h_2(K_1) = \frac{1}{2} h_2(k)h_2(d_2d_3).$$

Now we rewrite the formula for $h_2(L)$ in terms of $h_2(K_i)$. From above, notice that for example $h_2(d_2d_3) = 2h_2(K_1)/h_2(k)$, etc. and so by class field theory,

$$h_2(d_2d_3) = \frac{2[K_1 : K_1]}{[k : k]} = \frac{2[K_1 : k_1][K_1 : K_1]}{[k : k][K_1 : k]} = [K_1 : k_1].$$

Substituting into the above formula yields

$$[L_1 : L] = \frac{1}{32} q(L/\mathbb{Q})[k^1 : k][K_1^1 : k][K_1^2 : k][K_1^3 : k]$$

and since $[L : k] = 4$, we have

$$[L_1 : k_1] = \frac{1}{2} q(L/\mathbb{Q})[K_1^1 : k][K_2^1 : k][K_3^1 : k].$$

Notice, in particular, that if the 2-class field tower of $k$ is of length 1, then all the field degrees in the above formula equal 1, and therefore $q = 8$. Now, the length of the 2-class field tower of $k$ is 1 precisely when $d_i < 0$ for $i = 1, 2, 3$; see for example [Benjamin et al. 1997]. From this we have $8 = q(L/\mathbb{Q}) = Q(L/K)q(K/\mathbb{Q})$ from which it follows (by the comments before the proposition) that $Q(L/K) = 2$ and $q(K/\mathbb{Q}) = 4.$
Now suppose that \( d_1, d_2 > 0, d_3 < 0 \). In this case we have \( q(L/\mathbb{Q}) = q(K/\mathbb{Q}) \), where \( K = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}) \). Kuroda’s class number formula implies
\[
h_2(K) = \frac{1}{4}q(K/\mathbb{Q})h_2(F),
\]
where \( F = \mathbb{Q}(\sqrt{d_1d_2}) \). Then notice that \( \text{Cl}_2(F) \) is cyclic and thus \( F^1 = F^2 \). Thus since \( K/F \) is unramified so \( K \subseteq F^1 \), \( h_2(K) = h_2(F)/2 \). Plugging this into the formula above yields \( q(K/\mathbb{Q}) = 2 \). Thus \( q(L/\mathbb{Q}) = 2 \). □

**Proposition 3.** Let \( k \) be a complex quadratic field with 2-rank \( \text{Cl}(k) \leq 2 \) and with 4-rank \( \text{Cl}(k) \leq 1 \). Then
\[
\prod_i K_i^1 = \left( \prod_i K_i \right)^1,
\]
where \( K_i \) range over all the unramified quadratic extensions of \( k \).

**Proof.** If \( k^1 = k^2 \), then the proposition is trivially true, since both fields are \( k^1 \).
Thus, assume \( k^1 \neq k^2 \). Hence we know \( k = \mathbb{Q}(\sqrt{d_1d_2d_3}) \) where \( d_1, d_2 > 0, d_3 < 0 \).

From the proof of Theorem 2,
\[
[L^1 : k^1] = \frac{1}{4}[K^1_1 : k^1][K^1_2 : k^1][K^1_3 : k^1],
\]
where \( L = K_1K_2K_3 \) with \( K_i = k(\sqrt{d_i}) \). But notice that
\[
[K^1_1K^1_2K^1_3 : k^1] = \frac{[K^1_1 : k^1]}{[K^1_1 \cap K^1_2K^1_3 : k^1] \cdot [K^1_2 : k^1]} \cdot [K^1_3 : k^1].
\]
(Also notice this equation is true for any permutation of the indices.)

Now since
\[
[L^1 : k^1] = [L^1 : K^1_1K^1_2K^1_3][K^1_1K^1_2K^1_3 : k^1],
\]
we see by putting these equations together that
\[
[L^1 : K^1_1K^1_2K^1_3] = \frac{1}{4}[K^1_1 \cap K^1_2K^1_3 : k^1][K^1_2 \cap K^1_3 : k^1][K^1_3 \cap K^1_1 : k^1].
\]
To finish the proof, it suffices to show that
\[
[K^1_1 \cap K^1_2K^1_3 : k^1] = [K^1_2 \cap K^1_3 : k^1] = 2.
\]
Here is where some group theory comes in. Let \( G = \text{Gal}(k^2/k) \), and further let \( H_1, H_2, H_3 \) be the three maximal subgroups of \( G \) such that \( \text{Gal}(k^2/K_i) = H_i \). Then we need to show that
\[
(G' : H'_1H'_2) = (G' : H'_1(H'_2 \cap H'_3)) = 2.
\]
Here is a sketch of the proof. If \( G' \) is cyclic, say \( G' = \langle c \rangle \), by the table of possible groups and their presentations at the end of [Benjamin et al. 1997], we have (without loss of generality) \( H'_3 = \langle c^2 \rangle \) and \( H'_1H'_2 = \langle c^2 \rangle \), from which our result follows.
Now suppose $G'$ is not cyclic. Then by our assumption on the class group of $k$, $G$ must be nonmetacyclic with $G/G' \simeq (2, 2^n)$ for some $n > 1$. Now we assume the notation before [Benjamin et al. 2001, Lemma 1]. Hence let $G = \langle a, b \rangle$ where $a^2 \equiv b^{2^m} \equiv 1 \mod G'$. Let $[a, b] = c$ and define inductively, $c_2 = c$ and $c_{j+1} = [b, c_j]$. We have $G' = \langle c_2, c_3, \ldots \rangle$, and $G_3 = \langle c_2^2, c_3, \ldots \rangle$, and $G_4 = \langle c_2^2, c_3^2, c_4, \ldots \rangle$; see [Benjamin et al. 1997, Lemma 2]. Now if $H_3 = \langle a, b^2, G' \rangle$, then it is easy to see that $H_3^2G_4 = G_3$. Thus $H_3^2 = G_3$ by [Hall 1933, Theorem 2.49ii]. Hence $(G' : H_3^2) = (G' : G_3) = 2$. Similarly, if $H_1 = \langle b, G' \rangle$ and $H_2 = \langle ab, G' \rangle$, then $H_1H_2G_4 = G_3$ so once again $H_1H_2^2 = G_3$. This shows the result and finishes the proof of the proposition.

Now we consider $Cl_2(k) \simeq (2, 2, 2)$, and thus in particular disc $k = d_1d_2d_3d_4$ for distinct prime discriminants $d_i$. If we assume the $\Lambda$-rank of $Cl_2(k^1/k)$ is 2, then by [Benjamin et al. 2003, Theorem 2], exactly three of the $d_i$'s must be negative, say $d_1, d_2, d_3 < 0, d_4 > 0$.

**Theorem 4.** Let $k$ be a complex quadratic field with $Cl_2(k) \simeq (2, 2, 2)$. If the $\Lambda$-rank of $Cl_2(k^1)$ equals 2, the unit index $q(k^1/Q)$ equals $2^7$.

**Proof.** If $Cl_2(k) \simeq (2, 2, 2)$, then $Cl_2(k^1/k)$ has $\Lambda$-rank 2 if and only if $G/G' \simeq (2, 2, 2)$ and $G'/G_3 \simeq (2, 2)$, where $G = \text{Gal}(k^2/k)$. Thus $(G : G_3) = 32$ and $G'/G_3 \simeq (2, 2)$, and by [Hall and Senior 1964], $G/G_3$ must be one of the seven groups 32.033, 32.035, 32.036, 32.037, 32.038, 32.040, 32.041, in the notation of that same reference.

Let $L = k^1 = k_{\text{gen}}$. Kuroda’s class number formula (with $t = 4$, so $v = 16$) gives

\begin{equation}
(1) \quad h_2(L) = \frac{1}{2^{16}} q(L/Q) h_2(k) \prod_i h_2(k_i),
\end{equation}

where the $k_i$ are the quadratic subfields of $L$ excluding $k$.

The following table lists the 2-class numbers $h_2(k_i)$ and $h_2(L)$:

<table>
<thead>
<tr>
<th>$G/G_3$</th>
<th>$h_2(k_i)$</th>
<th>$h_2(L)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>32.041, 32.040</td>
<td>1 (7), 2 (6), 4</td>
<td>4</td>
</tr>
<tr>
<td>32.035, 32.037, 32.038</td>
<td>1 (7), 2 (5), 4, $2^n$</td>
<td>$2^n+1$</td>
</tr>
<tr>
<td>32.036</td>
<td>1 (7), 2 (5), $2^{m+1}$, $2^n$</td>
<td>$2^{m+n}$</td>
</tr>
<tr>
<td>32.033</td>
<td>1 (7), 2 (3), 4, $2^l$, $2^n$, $2^n$</td>
<td>$2^{l+m+n+1}$</td>
</tr>
</tbody>
</table>

Here “1 (7)” means that 7 quadratic subfields have 2-class number equal to 1. Plugging these data into (1) we immediately find the values of the unit index $q(L/Q)$ in each of the cases.

The 2-class numbers of the quadratic subfields $k_i$ of $L$ are easily determined using genus theory (see [Kaplan 1976], for instance). The 2-class numbers of $L$
were computed in [Benjamin et al. 2003], except for the group \( G/G_3 = 32.033 \) (for the first five groups, we have given the structure of \( G' \) explicitly; in the case \( G/G_3 = 32.036 \), we computed the 2-class number and actually showed that \( q(L/\mathbb{Q}) = 2^7 \)).

We will now study the case \( G/G_3 \simeq 32.033 \) in detail. By Proposition 16 in the same reference, we have \( k = \mathbb{Q}(\sqrt{d_1d_2d_3d_4}) \) with \( d_i < 0, (i = 1, 2, 3), d_4 > 0 \) and such that

\[
\left( \frac{d_1}{p_2} \right) = \left( \frac{d_2}{p_3} \right) = \left( \frac{d_3}{p_1} \right) = \left( \frac{d_4}{p_4} \right) = -1, \quad \left( \frac{d_4}{p_2} \right) = \left( \frac{d_4}{p_3} \right) = +1.
\]

Here is a list of the 2-class numbers of the quadratic subfields of \( L = k^1 \). Along with \( h_2(k) = 8 \), we have

\[
h_2(d_j) = h_2(d_1d_2) = h_2(d_2d_3) = h_2(d_1d_3) = 1, \quad (j = 1, 2, 3, 4)
\]
\[
h_2(d_1d_4) = h_2(d_1d_2d_4) = h_2(d_1d_3d_4) = 2, \quad h_2(d_1d_2d_3) = 4,
\]
\[
h_2(d_3d_4) = 2^l, \quad h_2(d_2d_3d_4) = 2^m, \quad h_2(d_2d_4) = 2^n \quad (l, m, n \geq 2).
\]

Let \( K = L^+ = \mathbb{Q}(\sqrt{d_1d_2}, \sqrt{d_1d_3}, \sqrt{d_1d_4}) \), the maximal real subfield of \( L \). Then \( q(L/\mathbb{Q}) = q(L/K)q(K/\mathbb{Q}) \) by [Benjamin et al. 2003, Proposition 1]. By [Lemmermeyer 1995], \( Q(L/K) = 2 \). In fact, if \( w_L (= #W_L) \equiv 2 \mod 4 \), then \( L = K(\sqrt{d_1}) \) and \( (p_1) = (\pi)^2 \) in \( Q(\sqrt{d_1d_2}) \) since \( p_1 \) ramifies and the field has odd class number. But then \( d_1 \mathcal{O}_K = (\pi \mathcal{O}_K)^2 \), and part (i)2(a) of [Lemmermeyer 1995, Theorem 1] implies \( Q(L/K) = 2 \). If instead \( w_L \equiv 4 \mod 8 \), then \( 2\mathcal{O}_K = (1+i)^2\mathcal{O}_K \), whence part (ii)2(a) of the same theorem shows again that \( Q(L/K) = 2 \).

Now we compute \( q(K/\mathbb{Q}) \). To this end, consider the quadratic number field \( k_0 = \mathbb{Q}(\sqrt{d_2d_3d_4}) \) with 2-class group \( \text{Cl}_2(k_0) = (2^m) \) and fundamental unit \( \varepsilon_{234} \). Then \( K/k_0 \) is a \( V_4 \)-extension with the quadratic subextensions \( K_1 = k_0(\sqrt{d_1d_2}), K_2 = k_0(\sqrt{d_1d_3}), K_3 = k_0(\sqrt{d_1d_4}) \). Let \( \varepsilon_{ij} \) denote the fundamental unit of \( \mathbb{Q}(\sqrt{d_id_j}) \) for \( 1 \leq i < j \leq 3 \). We shall determine \( \text{Cl}_2(K_1) \) and \( q(K_1/\mathbb{Q}) \). Since \( k_0 \) has cyclic 2-class group of type \( (2^m) \) and since \( K_1/k_0 \) is ramified, its class group contains \( (2^m) \) as a subgroup. If we can show that \( h_2(K_1) = 2^m \), then \( \text{Cl}_2(K_1) \simeq (2^m) \); since \( K/K_1 \) is unramified, it would then follow that \( \text{Cl}_2(K) \simeq (2^m-1) \). Applying Kuroda’s class number formula to \( K/\mathbb{Q} \) would then give \( q(K/\mathbb{Q}) = 2^6 \), and this in turn implies \( q(L/\mathbb{Q}) = 2^7 \) and \( h_2(L) = 2^{l+m+n-1} \).

For computing the 2-class number of \( K_1 \) we use Kuroda’s formula

\[
h_2(K_1) = \frac{1}{4} q(K_1/\mathbb{Q})h_2(d_1d_2)d_2(d_1d_3d_4)d_2(d_2d_3d_4) = q(K_1/\mathbb{Q})2^{m-1}.
\]

It suffices to show that \( q(K_1/\mathbb{Q}) \leq 2 \) (which implies \( q(K_1/\mathbb{Q}) = 2 \) by the argument above).

We consider two cases: \( d_k := \text{disc } k \not\equiv 4 \mod 8 \) and \( d_k \equiv 4 \mod 8 \). Assume \( d_k \not\equiv 4 \mod 8 \). The prime ideal above \( d_1 \) in \( \mathbb{Q}(\sqrt{d_1d_2}) \) is principal; hence \( X^2 - d_1d_2y^2 = \)
±4d_1 is solvable, and so is \( d_1 x^2 - d_2 y^2 = -4 \) (the minus sign must occur since \( (d_1/p_2) = -1 \)). Then \( \eta = \frac{1}{2} (x \sqrt{d_1} + y \sqrt{d_2}) \) is a unit in \( F = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}) \); note that \( \eta^2 < 0 \) in \( \mathbb{Q}(\sqrt{d_1 d_2}) \) since otherwise \( \eta \in \mathbb{R} \cap F = \mathbb{Q}(\sqrt{d_1 d_2}) \). Therefore \( \eta^2 = -\epsilon_{12}^2 \); notice that \( u \) is odd since otherwise \( \sqrt{-1} \in \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}) \), a contradiction. Thus

\[
-d_1 \epsilon_{12} = (\sqrt{d_1} \eta \epsilon_{12} (1-u)/2)^2
\]

is a square in \( \mathbb{Q}(\sqrt{d_1 d_2}) \).

Next consider \( \mathbb{Q}(\sqrt{d_1 d_3 d_4}) \), along with the diophantine equations

\[
d_1 x^2 - d_3 d_4 y^2 = \pm 4, \quad d_3 x^2 - d_1 d_4 y^2 = \pm 4, \quad d_4 x^2 - d_1 d_3 y^2 = \pm 4,
\]

which are solvable if the prime above \( d_1, d_3, d_4 \), respectively, is principal. The first implies \( (d_1/p_4) = +1 \), which contradicts the assumptions. The last implies \( (d_4/p_1) = (d_4/p_3) \), which also leads to a contradiction. Thus the second equation must have a solution, and reduction mod \( p_3 \) shows that we must have \( d_3 x^2 - d_1 d_4 y^2 = -4 \). Thus \( -d_3 \epsilon_{134} \) is a square in \( \mathbb{Q}(\sqrt{d_1 d_3 d_4}) \). Hence none of \( \epsilon_{12}, \epsilon_{134}, \epsilon_{12} \epsilon_{134} \) can be squares in \( K_1 \). Therefore \( q(K_1/\mathbb{Q}) \leq 2 \), as desired.

Now suppose \( d_k \equiv 4 \mod 8 \). Then our assumptions imply that \( d_3 = -4 \) or \( d_2 = -4 \). First assume that \( d_3 = -4 \). Then the argument above shows that \( \epsilon_{12} = p_1 \kappa^2 \) for some \( \kappa \in \mathbb{Q}(\sqrt{d_1 d_2}) \). Now consider \( \mathbb{Q}(\sqrt{d_1 d_3 d_4}) = \mathbb{Q}(\sqrt{p_1 p_4}) \). Then by genus theory ([Lemmermeyer 2000, page 76]) there is a principal ideal \( (\alpha) \) in \( \mathbb{Q}(\sqrt{p_1 p_4}) \) different from \( \langle 1 \rangle \) and \( \langle \sqrt{p_1 p_4} \rangle \) which is a product of distinct ramified prime ideals. We now consider the possibilities. First notice that the prime ideals above \( p_1 \) and \( p_4 \) are not principal, since otherwise \( p_1 x^2 - p_4 y^2 = \pm 1 \) is solvable which cannot happen. Now assume that the prime ideal above \( 2 \) is principal, equal to say \( (\pi) \) with \( \pi = x + y \sqrt{p_1 p_4} \), for some \( x, y \in \mathbb{N} \). Then \( \pi^2/2 = \mu \) a positive unit in \( \mathbb{Q}(\sqrt{p_1 p_4}) \). Clearly \( \mu \) is not a square in \( \mathbb{Q}(\sqrt{p_1 p_4}) \) since otherwise \( \sqrt{2} \in \mathbb{Q}(\sqrt{p_1 p_4}) \), a contradiction. Hence \( \epsilon_{134} = 2 \kappa^2 \), for some \( \kappa \in \mathbb{Q}(\sqrt{p_1 p_4}) \). Similarly \( \epsilon_{134} \) could be of the form \( 2 p_1 \kappa^2 \) or \( 2 p_4 \kappa^2 \). But in all of these cases we see that none of \( \epsilon_{12}, \epsilon_{134}, \epsilon_{12} \epsilon_{134} \) can be squares in \( K_1 \). Once again we have \( q(K_1/\mathbb{Q}) \leq 2 \).
Finally, suppose \( d_2 = -4 \). The argument above shows \( \varepsilon_{134} = p_3 \kappa^2 \), for some \( \kappa \in \mathbb{Q}(\sqrt{d_3d_4}) \). Now consider \( \mathbb{Q}(\sqrt{d_1d_2}) = \mathbb{Q}(\sqrt{p_1}) \). Then arguing as above we see \( \varepsilon_{12} = 2\kappa^2 \) or \( \varepsilon_{12} = 2p_4 \kappa^2 \) for some \( \kappa \in \mathbb{Q}(\sqrt{p_1}) \). But again this implies \( q(K_1/\mathbb{Q}) \leq 2 \); whence the result is established. \( \square \)

Now we come to the case where \( \text{Cl}_2(k) \simeq (2, 2) \) but with \( \text{disc } k \) divisible by three positive prime discriminants, say \( \text{disc } k = d_1d_2d_3d_4 \) with \( d_i > 0 \) for \( i = 1, 2, 3 \) and \( d_4 < 0 \). Our results in this case will be far less complete since our knowledge of \( \text{Gal}(k^2/k) \) is much more spotty. But we now simplify things somewhat by reducing to the maximal real subfield of \( k^1 \). To this end, from now on, let \( L = k^1 \) and \( K = L^+ \) the maximal real subfield of \( L \). But then

\[
q(L/\mathbb{Q}) = q(K/\mathbb{Q}),
\]

because \( q(L/\mathbb{Q}) = q(L/K)q(K/\mathbb{Q}) \) (by [Benjamin et al. 2003, Proposition 1], for example). By [Lemmermeyer 1995, Theorem 1] we get \( Q(L/K) = 1 \) since \( L = K(\sqrt{d_4}) \) is essentially ramified if \( d_4 \neq -4 \) and \( 2\mathbb{Q}_k \) is not an ideal square when \( d_4 = -4 \).

Now we need only consider \( K = \mathbb{Q}(\sqrt{p_1}, \sqrt{p_2}, \sqrt{p_3}) \) where \( p_i \) are the rational primes dividing \( d_i \). We set up the following notation. Let \( k_0 = \mathbb{Q}(\sqrt{p_1p_2p_3}) \). Let \( K_i = k_0(\sqrt{p_i}) \) for \( i = 1, 2, 3 \) and let \( k_i \) be the quadratic subfield of \( K_i \) not equal to \( k_0 \) and \( \mathbb{Q}(\sqrt{p_i}) \). (Notice that \( k_i = \mathbb{Q}(\sqrt{\text{disc } k_0/p_i}) \).) We now let \( \varepsilon_i \) for \( i = 0, 1, 2, 3 \) be the fundamental unit \( > 1 \) in \( k_i \) and \( N\varepsilon_i \) the norm from \( k_i \) to \( \mathbb{Q} \); also let \( \varepsilon_{p_i} \) be the fundamental unit in \( \mathbb{Q}(\sqrt{p_i}) \). Finally let \( H_i = \text{Gal}(k_i^2/k_i) \).

Now we assume that \( k_0 \) is a particular type of field. Namely, assume that \( \text{Cl}_2(k_0) \simeq (2, 2) \). This assumption implies that \( G = \text{Gal}(k_0^2/k_0) \) is one of the following types: abelian, quaternion, dihedral, semidihedral. Moreover notice that in this case \( k_0^2 = K = \mathbb{Q}(\sqrt{p_1}, \sqrt{p_2}, \sqrt{p_3}) \) since \( K/k_0 \) is unramified and \( h_2(k_0) = 4 \). Without loss of generality we now pick \( K_1 \) above so that \( H_1 \) is cyclic. We are in a position to state and prove the following (rather technical) theorem.

**Theorem 5.** Let \( k \) be a complex quadratic field with \( \text{Cl}_2(k) \simeq (2, 2) \) and with \( \text{disc } k = d_1d_2d_3d_4 \) where \( d_i \) are distinct prime discriminants divisible by primes \( p_i \) and \( d_1, d_2, d_3 \) are positive. With the notation above, assume that \( \text{Cl}_2(k_0) \simeq (2, 2) \). Then \( q = q(k^1/\mathbb{Q}) \) takes on the two values \( 2^4 \) and \( 2^5 \) as follows:

- If \( G \) is abelian, then \( q = 2^4 \).
- If \( G \) is nonabelian, then \( N\varepsilon_0 = +1 \) implies \( q = \begin{cases} 2^4 & \text{if } N\varepsilon_1 = -1, \\ 2^5 & \text{otherwise}, \end{cases} \), while

\[
N\varepsilon_0 = -1 \text{ implies } q = \begin{cases} 2^4 & \text{if } \left(N\varepsilon_1 = 1 \text{ or } (N\varepsilon_1 = -1 \text{ and } \sqrt{p_1\varepsilon_0\varepsilon_1} \notin K_1)\right) \\
&\quad \text{and } \left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_3}\right) = -1, \\
2^5 & \text{otherwise}. \end{cases}
\]
Proof: Since \( H_1 \) is cyclic (and so in particular abelian), we have \( k_2^2 = K^1_1 \). Thus \( h_2(K_1) = [k_0^2 : k_0^1][k_0^1 : K_1] = 2h_2(k_0^1) \), and hence
\[
(*) \quad h_2(k_0^1) = \frac{1}{2} h_2(K_1).
\]

Next notice by Kuroda’s class number formula that
\[
(**) \quad h_2(k_0^1) = \frac{1}{2^7} q h_2(k_1)h_2(k_2)h_2(k_3),
\]
where we have used \( v = 9 \) and \( h_2(k_0) = 4 \). Now Kuroda’s class number formula for \( K_i \) yields
\[
(***) \quad h_2(K_i) = \frac{1}{4} q_i h_2(p_i)h_2(k_i)h_2(k_0) = q_i h_2(k_i).
\]

But then \((*)\), \((**)\), \((***)\) imply
\[
\frac{1}{2} q_1h_2(k_1) = \frac{1}{2^7} q h_2(k_1)\frac{h_2(K_2)}{q_2}\frac{h_2(K_3)}{q_3},
\]
and therefore
\[
q = \frac{2^6 q_1q_2q_3}{h_2(K_2)h_2(K_3)}.
\]

Suppose first of all that \( G \) is abelian. Then \( k_0^2 = k_0^1 \) and so \( h_2(K_i) = 2 \). Thus \( q = 2^4 q_1q_2q_3 \). But \( h_2(k_i) \equiv 0 \mod 2 \), since the \( d_j > 0 \) for \( j = 1, 2, 3 \). So \((***)\) implies that \( h_2(k_i) = 2 = q_i h_2(k_i) \) and this in turn yields \( h_2(k_i) = 2 \) and \( q_i = 1 \), for \( i = 1, 2, 3 \). Thus when \( G \) is abelian, \( q = 2^4 \).

Now assume that \( G \) is not abelian. Then for \( G \simeq H_8 \), the quaternion group of order 8, or \( G \simeq D_4 \), the dihedral group of order 8, \( H_i \) has order 4 for \( i = 1, 2, 3 \) and so in particular \( h_2(K_i) = 4 \). If \( G \not\simeq H_8 \) or \( D_4 \), then \( H_2, H_3 \) are either dihedral, semidihedral, or quaternion, whence in particular the abelianizations \( H_2^{ab} \cong H_2^{ab} \cong (2, 2) \) and thus \( h_2(K_2) = h_2(K_3) = 4 \). Then
\[
q = 2^2 q_1q_2q_3.
\]

Case 1. Assume \( N \varepsilon_0 = 1 \). We now compute the \( q_i \)’s. First consider \( q_2 \). By [Couture and Derhem 1992, Theorem 1], \( (p_1/p_3) = -1 \), whence \( h_2(k_2)(= h_2(p_1p_3)) = 2 \).
But \( 4 = h_2(K_2) = q_2h_2(k_2) = 2q_2 \), so thus
\[
q_2 = 2.
\]

Next consider \( q_3 \). Again by [Couture and Derhem 1992, Theorem 1], \( (p_2/p_1) = 1 \) and \( (p_1/p_2)^4 = -(p_2/p_1)^4 \). Since \( 4 = h_2(K_3) = q_3h_2(p_1p_2) \), then either \( (h_2(p_1p_2) = 2 \text{ & } q_3 = 2) \) or \( (h_2(p_1p_2) = 4 \text{ & } q_3 = 1) \). We claim the latter does not hold. For, first by (\( \alpha \)) on page 318 of [Kaplan 1976], \( \text{Cl}_2^+(k_3) \simeq (4) \).
Hence if the latter holds, then \( N \varepsilon_3 = -1 \), which is not possible by [Kaplan 1976, Corollary 1]. Hence

\[ q_3 = 2. \]

Finally consider \( q_1 \). First assume \( N \varepsilon_1 = -1 \). Then by [Kubota 1956] the only possible square root of a nonsquare unit in \( K_1 \) would be \( \sqrt{\varepsilon_0} \) since the others have negative norm. Now applying [Benjamin et al. 1998, Proposition 3], for example, we see \( k_0(\sqrt{\varepsilon_0}) = k_0(\sqrt{\delta}) \), where by genus theory \( \delta \mid p_1 p_2 p_3 \) (but \( \neq \)) and \( \chi_j(\delta) = 1 \) for all genus characters of \( k_0 \). But then since \( (p_1/p_2) = (p_3/p_2) = 1 \), \( p_2 \) is trivial for all the genus characters and no other \( p_i \) has this property. Thus we may assume \( \delta = p_2 \) which is not a square in \( K_1 \). Thus

\[ N \varepsilon_1 = -1 \quad \text{implies} \quad q_1 = 1. \]

Now assume \( N \varepsilon_1 = +1 \). Then \( \delta_{k_0} = p_2 \) again and \( \delta_{k_1} = p_2 \) so that \( \varepsilon_0 \varepsilon_1 \) is a square in \( K_1 \) this time; again see [Kubota 1956]. Hence

\[ N \varepsilon_1 = +1 \quad \text{implies} \quad q_1 = 2. \]

Therefore, for \( N \varepsilon_0 = 1 \), \( q = 2^4 \) if \( N \varepsilon_1 = -1 \) and \( q = 2^5 \) if \( N \varepsilon_1 = +1 \).

Case 2. Assume \( N \varepsilon_0 = -1 \). Since \( \text{Cl}_2(k_0) = \text{Cl}_2^+(k_0) \) is elementary, the Rédei–Reichardt conditions [1933] imply that

\( a) \quad \left( \frac{p_i}{p_j} \right) = -1, \ \text{for all} \ i \neq j, \ \text{or} \ b) \ \left( \frac{p_1}{p_2} \right) = \left( \frac{p_1}{p_3} \right) = -\left( \frac{p_2}{p_3} \right) = -1. \)

First consider \( a) \). Then \( (p_i p_j/p_\ell) = 1 \) for all distinct \( i, j, \ell = 1, 2, 3 \). By [Couture and Derhem 1992, Theorem 2], \( G \simeq (2, 2) \) or \( H_8 \). Hence in the present situation \( G \simeq H_8 \). Thus as noted above the order of \( H_i \) is 4 whence \( h_2(K_i) = 4 \) so that \( 4 = h_2(K_i) = q_i h_2(k_i) \), for \( i = 1, 2, 3 \). But \( (p_i/p_j) = -1 \) implies \( h_2(k_i) = 2 \). Therefore, \( q_i = 2 \) for \( i = 1, 2, 3 \) so \( q = 2^5 \).

Next consider \( b) \). As immediately above, \( q_2 = q_3 = 2 \). Now consider \( q_1 \). If \( N \varepsilon_1 = +1 \), then arguing as above shows \( q_1 = 1 \). If \( N \varepsilon_1 = -1 \), so that the norms of \( \varepsilon_{p_i}, e_1, e_0 \) are negative, then \( q_1 = 1 \) if \( \sqrt{\varepsilon_{p_1} e_1 e_0} \notin K_1 \), and \( q_1 = 2 \) otherwise.

This establishes the theorem.

As a corollary to this theorem, we see that the structure \( G = \text{Gal}(k_0^2/k_0) \) determines \( q(k^1/Q) \):

**Corollary 6.** Let \( k_0 \) satisfy all the conditions in Theorem 5. For \( G = \text{Gal}(k_0^2/k_0) \),

\[
q(k^1/Q) = \begin{cases} 
2^4 & \text{if } G \text{ is abelian or dihedral}, \\
2^5 & \text{if } G \text{ is semidihedral or quaternion}. 
\end{cases}
\]
Proof: This follows immediately by Theorem 1 of [Couture and Derhem 1992] and a stronger form of part of Theorem 2 of the same paper, as found in [Lemmermeyer 1994a]. The main change in Theorem 2 is the following: with the notation above Theorem 5 suppose \( \sqrt{\varepsilon_0} = N \varepsilon_1 = -1 \). If \( \sqrt{\varepsilon_1 \varepsilon_0} \in K_1 \), then \( G \) is quaternion (of order 8 or larger). If \( \sqrt{\varepsilon_1 \varepsilon_0} \notin K_1 \), then \( G \) is dihedral. \( \square \)

The previous theorem is a special case of the following proposition:

**Proposition 7.** Let \( k \) be a complex quadratic field with \( \text{Cl}_2(k) \cong (2, 2, 2) \) and with \( \text{disc } k = d_1 d_2 d_3 d_4 \) where \( d_i \) are distinct prime discriminants divisible by primes \( p_i \) and \( d_1, d_2, d_3 \) are positive. With the notation above, assume that \( k_0 = k_1 \). Then \( q = q(k^1/\mathbb{Q}) = 2^4 \).

Proof: Recall that \( k_0 = \mathbb{Q}(\sqrt{p_1 p_2 p_3}) \), \( K = \mathbb{Q}(\sqrt{p_1}, \sqrt{p_2}, \sqrt{p_3}) \), and that \( q = q(K/\mathbb{Q}) \). Then Kuroda’s class number formula yields

\[
h_2(K) = \frac{1}{29} q \ h_2(k_0) h_2(k_1) h_2(k_2) h_2(k_3);\]

(again, refer to the notation before the previous theorem). Now since \( k_0 \subseteq K \subseteq k_1 \), we see \( k_0 \subseteq K := k_0 \subseteq k_1 \); but since by assumption \( k_0 = k_1 \), we have \( K = k_1 \). Hence

\[
h_2(K) = [K^1 : K] = [k_0 : K] = \frac{1}{4} [k_0 : k_0] = \frac{1}{4} h_2(k_0).
\]

Similarly \( K_i = k_0 \), for \( i = 1, 2, 3 \), whence

\[
h_2(K_i) = \frac{1}{2} h_2(k_0).
\]

On the other hand Kuroda’s class number formula again yields

\[
h_2(K_i) = \frac{1}{4} q_i \ h_2(k_0) h_2(k_i).
\]

All this implies

\[
\frac{1}{2} h_2(k_0) = \frac{1}{4} q_i \ h_2(k_0) h_2(k_i)
\]

so that \( 2 = q_i \ h_2(k_i) \). But then since \( 2 \mid h_2(k_i) \), we must have \( h_2(k_i) = 2 \). But then from above we have

\[
\frac{1}{4} h_2(k_0) = h_2(K) = \frac{1}{29} q \ h_2(k_0) h_2(k_1) h_2(k_2) h_2(k_3) = \frac{1}{26} q \ h_2(k_0).
\]

Therefore by solving for \( q \), we obtain

\[
q = 2^4.
\]

\( \square \)
3. Examples

We now give numerical examples illustrating Theorem 5 with \( q = 2^4 \) and \( q = 2^5 \).

**Example 1.** Let \( k_0 = \mathbb{Q}(\sqrt{2405}) = \mathbb{Q}(\sqrt{5 \cdot 13 \cdot 37}) \) and \( K = \mathbb{Q}(\sqrt{5}, \sqrt{13}, \sqrt{37}) \). By [Rédei and Reichardt 1933] or [Kaplan 1976] we see that \( \text{Cl}_2(k_0) \cong (2, 2) \). Moreover, we have \( N \varepsilon_0 = -1 \) and \( (13/5) = (37/5) = (37/13) = -1 \). Thus by [Couture and Derhem 1992, Theorem 2], \( \text{Gal}(k_0^2/k_0) \cong H_8 \) or \( (2, 2) \); but [Benjamin et al. 1998, Theorem 1] then shows that \( \text{Gal}(k_0^2/k_0) \cong H_8 \). Finally Theorem 4 above shows \( q = 2^5 \).

**Example 2.** Consider \( k_0 = \mathbb{Q}(\sqrt{290}) = \mathbb{Q}(\sqrt{2 \cdot 5 \cdot 29}) \); see the examples in [Couture and Derhem 1992]. Let \( K = \mathbb{Q}(\sqrt{2}, \sqrt{5}, \sqrt{29}) \). By [Rédei and Reichardt 1933] or [Kaplan 1976] or even [Couture and Derhem 1992], we see that \( \text{Cl}_2(k_0) \cong (2, 2) \). Moreover, we have \( N \varepsilon_0 = -1 \), where \( \varepsilon_0 = 17 + \sqrt{290} \) is the fundamental unit of \( k_0 \); and \( (2/5) = (2/29) = -(29/5) = -1 \). Now by genus theory \( K_1 = \mathbb{Q}(\sqrt{5 \cdot 29}, \sqrt{2}) \) (notation as in above). Also \( N \varepsilon_1 = -1 \) where \( \varepsilon_1 = 12 + \sqrt{145} \) is the fundamental unit of \( \mathbb{Q}(\sqrt{5 \cdot 29}) \). Finally, \( \varepsilon_2 = 1 + \sqrt{2} \). By the techniques described in [Kubota 1956] we see that \( \varepsilon_0 \varepsilon_1 \varepsilon_2 \) is not a square in \( K_1 \). Theorem 5 above then shows \( q = 2^4 \). Furthermore, [Couture and Derhem 1992, Theorem 2] and PARI show \( \text{Gal}(k_0^2/k_0) \cong D_4 \).

**References**


[Hall and Senior 1964] M. Hall, Jr. and J. K. Senior, *The groups of order \( 2^n \) \( (n \leq 6) \)*, Macmillan, New York, 1964. MR 29 #5889 Zbl 0192.11701


Received August 14, 2005.

ELLIOT BENJAMIN
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MAINE
ORONO, MAINE 04469
UNITED STATES
ben496@prexar.com

FRANZ LEMMERMEYER
DEPARTMENT OF MATHEMATICS
BILKENT UNIVERSITY
06800 BILKENT, ANKARA
TURKEY
franz@fen.bilkent.edu.tr

CHIP SNYDER
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MAINE
ORONO, MAINE 04469
UNITED STATES
snyder@math.umaine.edu
We define a class of symplectic Lie groups associated with solvable symmetric spaces. We give a universal strict deformation formula for every proper action of such a group on a smooth manifold. We define a functional space where performing an asymptotic expansion of the nonformal deformed product in powers of the deformation parameter yields an associative formal star product on the symplectic Lie group at hand. The cochains of the star product are explicitly given (without recursion) in the two-dimensional case of the affine group $ax + b$. The latter differs from the Giaquinto–Zhang construction, as shown by analyzing the invariance groups. In a Hopf algebra context, the above formal star product is shown to be a smash product and a compatible coproduct is constructed.

1. Introduction

The concept of universal deformation for abelian Lie group actions was introduced by Rieffel [1993] in the operator algebraic context. Later, the notion of a universal deformation formula (UDF) within a Hopf algebraic context was defined by Giaquinto and Zhang [1998] at the formal level. Here we study a similar notion within the framework of (nonabelian) solvable Lie group actions. Our construction of UDFs relies on both formal and nonformal aspects. The Lie groups considered here are symplectic semidirect products of abelian Lie groups, natural generalizations of the two-dimensional affine group $ax + b$. On each such group $\mathbb{S}$, we define a function space $\mathbb{A} \subset C^\infty(\mathbb{S})$ invariant under the left regular representation and endowed with a one-parameter family of products $\{\ast_\theta\}_{\theta \in \mathbb{R}}$ in such a way that each pair $(\mathbb{A}, \ast_\theta)$ is an associative (topological) algebra which the group $\mathbb{S}$ naturally acts on by automorphisms. As observed in [Bieliavsky 2002; Bieliavsky et al. 2003], such a data provides a UDF in the following way. Let $M$ denote

---

**MSC2000:** primary 54C40, 14E20; secondary 46E25, 20C20.

**Keywords:** universal deformation formula, symplectic Lie group, symmetric space, Lie group actions.

Maeda is partially supported by the 21st Century COE Program “Integrative Mathematical Sciences, Progress in Mathematics Motivated by Natural and Social Phenomena”, and by Grant-in-Aid for Scientific Research #18204006, Ministry of Education, Science and Culture, Japan.
a smooth manifold which the group $\mathbb{S}$ acts on by diffeomorphisms. Denote by $
abla : \mathbb{S} \times C^\infty(M) \to C^\infty(M) : (s, u) \mapsto \nabla_s(u)$ the induced action at the level of functions and set $\nabla^s(u)(x) := \nabla_s(u)(x)$ for $u \in C^\infty(M)$, $x \in M$ and $s \in \mathbb{S}$. Then, the function space $\mathbb{B} := \{u \in C^\infty(M) \mid \nabla^s(u)(x) \in \mathbb{A}_\theta \text{ for all } x \in M\}$ is naturally endowed with a one-parameter family of associative products $\{\star^\theta_M\}$ defined by the formula $u \star^\theta_M v(x) := (\nabla^x(u) \star^\theta \nabla^x(v))(e)$ ($e$ denotes the identity element in $\mathbb{S}$). The problem is of course to show that the space $\mathbb{B}$ is sufficiently rich, in the sense that it contains at least the smooth compactly supported functions on $M$. This is the case for the class of groups $\mathbb{S}$ considered here provided the action of $\mathbb{S}$ on $M$ is proper. Next, comes the question of defining an appropriate functional framework allowing to pass from our nonformal setting to the formal framework of star products [Bayen et al. 1978a; 1978b]. The above-mentioned nonformal universal deformation formulae are of the oscillatory — or WKB — type. This means that the product $a \star^\theta b$ is defined by an integral expression $a \star^\theta b = \int_{\mathbb{S} \times \mathbb{S}} K_\theta a \otimes b$ where the kernel $K_\theta$ has the oscillatory form $K_\theta = \theta^m a_\theta e^{i/\theta}S$, where $a_\theta$ and $S$ belong to $C^\infty(\mathbb{S} \times \mathbb{S} \times \mathbb{S}, \mathbb{R})$. In particular, for $a$ and $b$ smooth compactly supported, one may perform a stationary phase expansion of $a \star^\theta b$ in powers of $\theta$ yielding a formal product on $C^\infty(\mathbb{S})[\theta]$. The property of associativity of this formal product depends on functional properties of the nonformal algebras $(\mathbb{A}, \star^\theta)$.

In Section 2, we prove that every exact symplectic semidirect product $\mathbb{S}$ of two abelian groups always acts strictly transitively on an elementary solvable symplectic symmetric space in the sense of [Beliakovski 2002]. Conversely, we show that when complex every such space gives rise to an exact symplectic semidirect product of two abelian groups. This allows to identify in a $\mathbb{S}$-equivariant manner the manifolds underlying the group $\mathbb{S}$ and the corresponding symmetric space.

In Section 3, we observe that Section 2 together with the construction in [Beliakovski 2002] of nonformal quantizations on elementary solvable symplectic symmetric spaces yield on every such semidirect product $\mathbb{S}$ a left-invariant nonformal deformation quantization. As an immediate consequence, the latter gives rise to strict deformation quantizations for the proper actions of $\mathbb{S}$ on smooth manifolds.

In Section 4, restricting to the case dim $\mathbb{S} = 2$, we associate to our nonformal deformation an associative left-invariant formal star product on $C^\infty(\mathbb{S})[\theta]$ for which we give the cochains totally explicitly (without any recursion). This is essentially done by defining a $\theta$-independent functional space (denoted hereafter $\mathcal{E}$) closed under our one-parameter family of nonformal deformations $\{\cdot^\theta\}$ and by studying the smoothness (in a suitable sense) of our algebras $(\mathcal{E}, \cdot^\theta)$. For nonzero values of $\theta$, the space $\mathcal{E}$ plays an analogous role as the Schwartz space does in the case of Weyl’s quantization. As it clearly appears, the restriction to the dimension two is inessential.
In Section 5, we show that the above mentioned formal star product can be seen as a smash product of $C^\infty(\mathbb{R})[\theta]$ and the Hopf algebra of polynomials on $\mathbb{R}$. From the latter, we deduce a compatible formal coproduct and a Hopf structure on $C^\infty(\mathbb{S})[\theta]$.

2. Symplectic Lie algebras associated to a class of symmetric spaces

Definition 2.1. Following the terminology of [Lichnerowicz and Medina 1988], a symplectic Lie algebra is a pair $(\mathfrak{s}, \omega)$ where $\mathfrak{s}$ is a Lie algebra and $\omega \in \wedge^2(\mathfrak{s}^*)$ is a nondegenerate Chevalley two-cocycle with respect to the trivial representation of $\mathfrak{s}$.

In this section, we associate symplectic Lie algebras to a class of (infinitesimal) symplectic symmetric spaces.

Definition 2.2 [Bieliavsky et al. 1995; Bieliavsky 1995]. Let $(\mathfrak{g}, \sigma)$ be an involutive algebra, meaning that $\mathfrak{g}$ is a finite dimensional real Lie algebra and $\sigma$ is an involutive automorphism of $\mathfrak{g}$. Let $\Omega$ be a skewsymmetric bilinear form on $\mathfrak{g}$. Then the triple $(\mathfrak{g}, \sigma, \Omega)$ is called a symplectic triple if the following properties are satisfied.

(i) Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ where $\mathfrak{k}$ (resp. $\mathfrak{p}$) is the $+1$ (resp. $-1$) eigenspace of $\sigma$. Then $[\mathfrak{p}, \mathfrak{p}] = \mathfrak{k}$ and the representation of $\mathfrak{k}$ on $\mathfrak{p}$, given by the adjoint action, is faithful.

(ii) $\Omega$ is a Chevalley 2-cocycle for the trivial representation of $\mathfrak{g}$ on $\mathbb{R}$ such that for any $X$ in $\mathfrak{k}$, $i(X)\Omega = 0$. Moreover, the restriction of $\Omega$ to $\mathfrak{p} \times \mathfrak{p}$ is nondegenerate.

Two such triples $(\mathfrak{g}_i, \sigma_i, \Omega_i) (i = 1, 2)$ are isomorphic if there exists a Lie algebra isomorphism $\psi : \mathfrak{g}_1 \rightarrow \mathfrak{g}_2$ such that $\psi \circ \sigma_1 = \sigma_2 \circ \psi$ and $\psi^*\Omega_2 = \Omega_1$.

Such a triple is called indecomposable if it cannot be expressed as a direct sum of symplectic triples.

Definition 2.3. A symplectic triple $t = (\mathfrak{g}, \sigma, \Omega)$ is called holonomy isotropic, or HI, if $[\mathfrak{g}, \mathfrak{g}]$ is an isotropic subspace of $(\mathfrak{p}, \Omega)$.

Proposition 2.1 [Bieliavsky 1998]. A symplectic triple $t = (\mathfrak{g}, \sigma, \Omega)$ is holonomy isotropic if and only if $[\mathfrak{g}, \mathfrak{g}]$ is abelian.

Definition 2.4. Let $t = (\mathfrak{g}, \sigma, \Omega)$ be HI and consider the extension sequence

$$0 \rightarrow [\mathfrak{g}, \mathfrak{g}] \rightarrow \mathfrak{g} \rightarrow \mathfrak{a} := \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}] \rightarrow 0.$$

The HI triple $t$ is called split if this extension is split.

Lemma 2.2. Let $t = (\mathfrak{g}, \sigma, \Omega)$ be HI split. Set $b = [\mathfrak{g}, \mathfrak{g}]$ and denote by $\rho : \mathfrak{a} \rightarrow \text{End}(b)$ the splitting homomorphism. Then, realizing $\mathfrak{g}$ as the semidirect product $\mathfrak{g} = b \times_\rho \mathfrak{a}$, one can assume that $\mathfrak{a}$ is stable under $\sigma$. 
Proof: For \( a \in a \subset g \), write \( a = a_t + a_p \) according to the decomposition with respect to \( \sigma \). Then for all \( a, a' \in a \), one has \( 0 = [a, a'] = [a_p, a'_p] + b \quad b \in [\mathfrak{t}, \mathfrak{p}] \) since \( \mathfrak{t} \) is abelian. This yields \([a_p, a'_p] = 0\).

Therefore, for \( pr_p : g \rightarrow p \) the projection parallel to \( \mathfrak{t} \), the \( p \)-component \( pr_p(a) \) is an abelian subalgebra of \( g \) supplementary to \( b \). A dimension count then yields the lemma. \( \square \)

Lemma 2.3. Assume that \( t = (g, \sigma, \Omega) \) is HI split, indecomposable and nonflat. Set \( 0 \rightarrow b \rightarrow g \rightarrow a \rightarrow 0 \) as in Lemma 2.2. Then \( a \) and \( \mathfrak{l} = [\mathfrak{t}, \mathfrak{p}] \) are in duality. In particular, there exists a \( \mathfrak{t} \)-invariant symplectic structure on \( p \) for which \( a \) is Lagrangian.

Proof: Set \( V := \mathfrak{l}^\perp \cap a \) and choose a subspace \( W \) of \( a \) in duality with \( \mathfrak{l} \). Counting dimensions yields \( a = W \oplus V \). Moreover, in the decomposition \( p = \mathfrak{l} \oplus W \oplus V \), the matrix of \( \Omega \) is of the form

\[
[\Omega] = \begin{pmatrix}
0 & I & 0 \\
-I & 0 & B \\
0 & -B' & A
\end{pmatrix}.
\]

Since \( \det[\Omega] \neq 0 \), one gets \( \det\begin{pmatrix}
-I & B \\
0 & A
\end{pmatrix} \neq 0 \); hence \( \det A \neq 0 \) and \( V \) is symplectic.

Now, \( \Omega([\mathfrak{t}, V], p) = \Omega(V, \mathfrak{l}) = 0 \), hence \([\mathfrak{t}, V] = 0\). Also \([V, \mathfrak{l}] = [V, [\mathfrak{t}, p]] = 0\) by Jacobi. Thus \( V \) is central, and therefore trivial by indecomposability. \( \square \)

We now assume that \((g^1, \sigma^1)\) is the involutive Lie algebra underlying a split HI symplectic triple which is indecomposable and nonflat. We fix \( \Omega^1 \) such that the HI symplectic triple \( t^1 = (g^1, \sigma^1, \Omega^1) \) with \( 0 \rightarrow b^1 \rightarrow g^1 \rightarrow a^1 \rightarrow 0 \) has \( a^1 \) and \( \mathfrak{l}^1 = [\mathfrak{l}^1, \mathfrak{p}^1] \) dual Lagrangian subspaces. We then consider the associated exact triple [Beliavsky 1998], which we denote by \( t = (g, \sigma, \Omega) \) (if \( t^1 \) is already exact we set \( t = t^1 \)). Since \( a^1 \) is isotropic, the triple \( t \) is elementary solvable with

\[
0 \rightarrow b := [g, g] \rightarrow g \rightarrow a := a^1 \rightarrow 0.
\]

We now follow a procedure as in [Beliavsky 2002]. The map \( \rho : a \rightarrow \text{End}(b) \) is injective (because \( \Omega \) is nondegenerate), so we may identify \( a \) with its image \( : a = \rho(a) \). Let \( \Sigma : \text{End}(b) \rightarrow \text{End}(b) \) be the automorphism induced by the conjugation with respect to the involution \( \sigma|_b \in GL(b) \), i.e. \( \Sigma = Ad(\sigma|_b) \). The automorphism \( \Sigma \) is involutive and preserves the canonical Levi decomposition \( \text{End}(b) = \mathfrak{z} \oplus sl(b) \), where \( \mathfrak{z} \) denotes the center of \( \text{End}(b) \). Writing the element \( a = \rho(a) \in a \) as \( a = a_Z + a_0 \) with respect to this decomposition, one has \( \Sigma(a) = a_Z + \Sigma(a_0) = -a = -a_Z - a_0 \), because the endomorphisms \( a \) and \( \sigma|_b \) anticommute. Hence \( \Sigma(a_0) = -2a_Z - a_0 \) and therefore \( a_Z = 0 \). So, \( a \) actually lies in the semisimple part \( sl(b) \). For any \( x \in sl(b) \), we denote by \( x = x^5 + x^N \), \( x^5, x^N \in sl(b) \), its abstract Jordan–Chevalley decomposition. Observe that, for \( sl(b) = sl_+ \oplus sl_- \), the
decomposition in \((\pm 1)\Sigma\)-eigenspaces, one has \(a \subset sl_-\). Also, \(a_N := \{a^N\}_{a \in a}\) is an abelian subalgebra in \(sl_-\) commuting with \(a\). Set \(a_S := \{a^S\}_{a \in a}\).

Consider the complexification \(b^c := b \otimes \mathbb{C}\) and \(\mathbb{C}\)-linearly extend the endomorphisms \(\{\rho(a)\}_{a \in a}\) and \(\sigma\). Also consider the complex Lie algebra \(sl(b^c) = sl(b) \otimes \mathbb{C}\) and \(\mathbb{C}\)-linearly extend to \(sl(b^c)\) the involution \(\Sigma\).

Let
\[
\begin{equation}
(2) \quad b^c = \bigoplus_{\alpha \in \Phi} b_\alpha
\end{equation}
\]
be the weight space decomposition with respect to the action of \(a_S\). Note that for all \(\alpha\), one has \(a_N.b_\alpha \subset b_\alpha\). Moreover, for all \(X_\alpha \in b_\alpha\) and \(a^S \in a_S\), one has
\[
\sigma(a^S.X_\alpha) = \alpha(a^S)\sigma(X_\alpha) = \alpha a^S \sigma^{-1} \sigma X_\alpha = \Sigma(a^S) \sigma X_\alpha = -a^S \sigma(X_\alpha).
\]
Therefore, \(-\alpha \in \Phi\) and \(\sigma b_\alpha = b_{-\alpha}\). Note in particular that \(\sigma b_0 = b_0\).

**Lemma 2.4.** If the triple \(t^1\) is assumed indecomposable and nonflat, then
\[
b_0 = 0.
\]

**Proof.** Assume \(0 \in \Phi\). For all \(\alpha \in \Phi\), the subspace
\[
V_\alpha := b_\alpha \oplus b_{-\alpha}
\]
of \(b^c\) is stable under \(\sigma\). In particular, the complexified involutive Lie algebra \((g^c, \sigma)\), where \(g^c := g \otimes \mathbb{C}\), can be expressed as \(g^c = \mathfrak{a}^c \times_{\rho} b^c\) with
\[
b^c = \bigoplus_{\alpha \in \Phi^+} b_\alpha \oplus b_0,
\]
where the *positive system* of weights \(\Phi^+\) is chosen so that
\[
\Phi = \{0\} \cup \Phi^+ \cup (-\Phi^+)
\]
(disjoint union). One therefore has the decomposition
\[
V_\alpha = \mathfrak{k}_\alpha \oplus \mathfrak{l}_\alpha
\]
into \((\pm)\)-eigenspaces for \(\sigma\). Moreover, since \(g^c = [g^c, g^c] \) and \([a, b_\alpha] \subset b_\alpha\), one has
\[
\mathfrak{t}_\alpha = [a, \mathfrak{t}_\alpha] \quad \text{and} \quad \mathfrak{l}_\alpha = [a, \mathfrak{l}_\alpha],
\]
for all \(\alpha \in \Phi^+ \cup \{0\}\). This implies \(\mathfrak{l}_0 = [a_N, \mathfrak{t}_0] \) and \(\mathfrak{t}_0 = [a_N, \mathfrak{t}_0]\). Hence \(\mathfrak{l}_0 = [a_N, [a_N, \mathfrak{t}_0]]\) and an induction yields \(\mathfrak{l}_0 = 0\). \(\square\)

**Corollary 2.5.** A nilpotent HI split symplectic symmetric space is flat.
Proposition 2.6. Let $t = (g = b \times_\rho a, \sigma, \Omega = \delta \xi)$ be the exact triple associated with a nonflat indecomposable split symplectic triple $t^1$. Let $\Phi$ be the set of weights associated with the (complex) action of $a_{\mathbb{C}}$ on $b'$. Fix a positive system $\Phi^+$ and set

$$b^+ := \bigoplus_{\alpha \in \Phi^+} b_\alpha.$$ 

Then the pair $(s^c := a^c \times_\rho b^+, \Omega|_{s^c})$ is a (complex) symplectic Lie algebra.

Proof. By the proof of Lemma 2.4, the restricted projection $b^+ \overset{p}{\to} a$ mapping $X$ to $\frac{1}{2}(X - \sigma(X))$ is a linear isomorphism. Moreover, for all $X \in b^+, a \in a^c$, one has $\Omega(X, a) = \xi[p(c), a]$. The proposition follows from the nondegeneracy of the pairing $a^c \times a^c \to \mathbb{C}$. \qed

Definition 2.5. Let $t$ be an HI split symplectic triple. Decompose $t$ into a direct sum of indecomposables and a flat factor. Proposition 2.6 then canonically associates to $t$ a (complex) symplectic Lie algebra $s^c(t)$, the complex symplectic Lie algebra associated with $t$.

Definition 2.6. A symplectic Lie algebra $(s, \omega)$ is called elementary solvable if

(i) it is a split extension of abelian Lie algebras $a$ and $d$:

$$0 \to d \to s \to a \to 0;$$

(ii) The cocycle $\omega$ is exact.

Conversely to Proposition 2.6, one has

Proposition 2.7. Every elementary solvable symplectic Lie algebra is associated with a split HI symplectic symmetric space.

Proof. Denote by $\rho : a \to \text{End}(d)$ the splitting homomorphism and by $\overline{\rho} : a \to \text{End}(d)$ the opposite representation: $\overline{\rho}(a)(X) := -\rho(a)(X), \quad X \in d$. Set

$$b := d \oplus d$$

and let $a$ act on $b$ via $\rho \oplus \overline{\rho}$. Define the involution $\sigma_b$ of $b$ by

$$\sigma_b(X, Y) = (Y, X), \quad X, Y \in d.$$ 

Set

$$g := b \times_{\rho \oplus \overline{\rho}} a$$

and define the involution $\sigma$ of $g$ as

$$\sigma := \sigma_b \oplus (-\text{id}_d).$$

One then observes that $(g, \sigma)$ is an involutive Lie algebra. We have $\mathfrak{t} = \{(X, X)\}_{X \in d}$ while $\mathfrak{p} = \{(X, -X)\}_{X \in d}$. 
Let $\eta \in \mathfrak{d}^*$ be such that $\delta \eta = \omega$ and define $\xi \in \mathfrak{t}^*$ by

$$\xi(X, X) := \eta(X), \quad X \in \mathfrak{d}.$$  

Extending $\xi$ to $\mathfrak{g}$ by 0 on $\mathfrak{p}$, one defines a symplectic coboundary on $\mathfrak{g}$:

$$\Omega := \delta \xi.$$  

The triple $(\mathfrak{g}, \sigma, \Omega)$ then defines the desired elementary solvable symplectic symmetric space.


3. Strict deformation quantization for proper actions

We first recall that to an involutive Lie algebra, one associates a simply connected symmetric space $M$ in the usual way (see [Kobayashi and Nomizu 1969], for example). When associated with a symplectic triple, the space $M$ turns out to be naturally endowed with an invariant symplectic structure. The space $M$ is then called a symplectic symmetric space. One then has

**Theorem 3.1** [Bieliavsky 2002]. Let $M$ be the simply connected symplectic symmetric space associated with an HI split symplectic triple. Assume $M$ is strictly geodesically convex with respect to its canonical affine connexion. There is a one-parameter family $\{K_\theta \in C^\infty(M \times M \times M, \mathbb{C})\}_{\theta \in \mathbb{R}}$ of smooth invariant three-point kernels, and a corresponding family $\{\mathcal{H}_\theta \in C^\infty(M)\}_{\theta}$ of invariant function spaces, such that

(i) $\mathfrak{D}(M) := C^\infty_c(M) \subset \mathcal{H}_\theta$ for all $\theta$;

(ii) the formula

$$u \ast_\theta v(x) := \int_{M \times M} K(x, y, z) u(y) v(z) \, dy \, dz \quad (u, v \in \mathfrak{D}(M))$$  

extends to $\mathcal{H}_\theta$ as an associative $\mathbb{C}$-algebra product law denoted hereafter $\ast_\theta$, and the automorphism group of $M$ acts on the algebra $(\mathcal{H}_\theta, \ast_\theta)$ by algebra automorphisms;

(iii) for $u, v \in \mathfrak{D}(M)$ and $x \in M$, one has the asymptotic expansion

$$u \ast_\theta v(x) \sim uv(x) + \frac{\theta}{2i} \{u, v\}(x) + o(\theta^2),$$  

where $\{,\}$ denotes the invariant symplectic Poisson bracket on $C^\infty(M)$.

We now pass to the announced nonformal UDFs. In this section, $(\mathfrak{s}, \Omega)$ denotes an elementary solvable symplectic Lie algebra associated with a strictly geodesically convex HI split symplectic symmetric space $M$. We denote by $\mathfrak{S}$ the corresponding simply connected symplectic Lie group, whose identity element we write as $e$.  

Now consider a proper smooth action of $\mathbb{S}$ on a manifold $X$:
$$\mathbb{S} \times X \to X : (g, x) \mapsto \tau_g(x).$$
And denote by
$$\mathbb{S} \times C^\infty(X) \to C^\infty(X) : (g, a) \mapsto \alpha_g(a)$$
the corresponding left action on functions on $X$. At last, for all $a \in C^\infty(X)$ and $x \in X$, set
$$\alpha_x(a) : \mathbb{S} \to \mathbb{C} : g \mapsto \alpha_g(a)(x) := \tau_g \ast (a).$$

**Theorem 3.2.** Define the space
$$\mathcal{B}_\theta := \{ a \in C^\infty(X) \text{ such that } \alpha_x(a) \in \mathcal{H}_\theta \text{ for all } x \in X \}.$$
Then
(i) $\mathcal{D}(X) := C^\infty_c(X) \subset \mathcal{B}_\theta$ for all $\theta$, and
(ii) the formula
$$a \ast^X_{\theta} b(x) := (\alpha_x(a) \ast_{\theta} \alpha_x(b))(x)$$
defines an associative $\mathbb{C}$-algebra product law on the function space $\mathcal{B}_\theta$.

**Proof.** Item (i) follows from the properness of the action and item (i) of the preceding theorem. Item (ii) follows from [Beliavsky 2002, p. 282, item (ii)]; see also [Beliavsky et al. 2003]. \hfill $\square$

For more general actions, we shall remain at the formal level. We shall prove in the next section that the asymptotic expansion (5) actually defines an invariant formal star product on $M$. For simplicity, we shall treat only the case $\dim \mathbb{S} = 2$ but as it will appear this is not essential.

### 4. Formal universal deformation formulae

We present the $ax + b$ group as $\mathbb{S} = \mathbb{R}^2 = \{(a, \ell)\}$ with multiplication law
$$(a, \ell)(a', \ell') := (a + a', e^{-2a' \ell + \ell'}).$$
The Lie algebra $\mathfrak{s}$ of $\mathbb{S}$ is then realized as $\mathfrak{s} = \text{span}\{H, E\}$ with
$$(a, 0) := \exp(aH) \text{ and } (0, \ell) := \exp(\ell E).$$
For $X \in \mathfrak{g}$, we denote by $\tilde{X}$ the corresponding left-invariant vector field on $\mathbb{S}$.

We denote by $\mathcal{F}$ the Schwartz space on $\mathbb{S} = \{(a, \ell)\}$. For $u \in \mathcal{F}$, we denote by $\mathcal{F}(u)$ the partial Fourier transform
$$\mathcal{F}(u)(a, \xi) := \hat{u}(a, \xi) := \int e^{-i\xi \ell} u(a, \ell) d\ell.$$
We set $\mathbb{S} := \{(a, \xi)\}$ and we denote by $\tilde{\mathcal{F}}$ the space of Schwartz functions in these variables. Thus, one has the isomorphism $\mathcal{F} : \mathcal{S} \to \tilde{\mathcal{F}}$. For $\theta \in \mathbb{R}$, we define the following diffeomorphism $\phi_\theta : \mathbb{S} \to \mathbb{S}$:

$$\phi_\theta(a, \xi) := \begin{cases} (a, \frac{1}{2\theta} \sinh(2\theta \xi)) & \text{for } \theta \neq 0, \\ (a, \xi) & \text{for } \theta = 0. \end{cases}$$

Setting $\tau_\theta := \mathcal{F}^{-1} \circ (\phi_\theta^{-1})^* \circ \mathcal{F}$, we formally define the following distribution space on $\mathbb{S}$:

$$\mathcal{E}_\theta := \tau_\theta(\mathcal{F}).$$

**Lemma 4.1 [Beliavsky 2002].** For all $\theta \in \mathbb{R}$,

$$\mathcal{S} \subset \mathcal{E}_\theta \subset \mathcal{S}' \cap C^\infty(\mathbb{S});$$

where $\mathcal{S}'$ denotes the space of tempered distributions in variables $(a, \ell)$.

**Lemma 4.2.**

$$(\phi_\theta^{-1})^* (\tilde{\mathcal{F}}) = \begin{cases} (\phi_{1/2}^{-1})^* (\tilde{\mathcal{F}}) & \text{for } \theta \neq 0, \\ \tilde{\mathcal{F}} & \text{for } \theta = 0. \end{cases}$$

**Proof.** First observe that $f \in (\phi_\theta^{-1})^* \tilde{\mathcal{F}}$ if and only if for all $M, N \in \mathbb{N}$ one has $\int \xi^M \partial_\xi^N (\phi_\theta^* f) \, d\xi < \infty$. Let us restrict ourselves to the case $N = 1$, the general case being entirely similar. One has $\partial_\xi (\phi_\theta^* f)(\xi) = \cosh(2\theta \xi) \phi_\theta^* (\partial_\eta f) = \phi_\theta^* (\sqrt{1 + (2\theta \eta)^2} \partial_\eta f)$. Therefore, the condition becomes

$$\int \arcsinh(2\theta \eta) \partial_\eta f(\eta) \, d\eta < \infty.$$ 

Since the functions $\arcsinh(2\theta \eta) \partial_\eta f(\eta)$ and $\arcsinh \eta \partial_\eta f(\eta)$ have, up to a $\theta$-dependant multiple, the same asymptotic behavior for large $|\eta|$, the finiteness of the above integral is equivalent to the finiteness of $\int \arcsinh \eta \partial_\eta f(\eta) \, d\eta$. \hfill \Box

In particular, for all $\theta \in \mathbb{R}$, one has

$$(\phi_\theta^{-1})^* (\tilde{\mathcal{F}}) \subset (\phi_{1/2}^{-1})^* (\tilde{\mathcal{F}}).$$

The above lemma leads us to define the following spaces:

$$\hat{\mathcal{E}} := (\phi_{1/2}^{-1})^* (\tilde{\mathcal{F}}) \text{ and } \hat{\mathcal{E}} := \mathcal{F}^{-1}(\hat{\mathcal{E}}).$$

We then observe the following obvious fact.

**Lemma 4.3.** Denote by $\mathcal{B}_\infty(\hat{\mathbb{S}})$ the space of smooth functions $w \in C^\infty(\hat{\mathbb{S}})$ such that for all multi-index $\alpha$, $D^\alpha w \in C_0(\hat{\mathbb{S}})$. Then,

$$\hat{\mathcal{E}} \subset \mathcal{B}_\infty(\hat{\mathbb{S}}) \quad \text{and} \quad \partial_\xi^2(\hat{\mathcal{E}}) \subset L^1(\hat{\mathbb{S}}).$$
Lemma 4.4. Let \( u \in \mathcal{F} \). Then,

(i) there exists a dense open subset \( \Omega \subset G \) where, for all \( \theta \in \mathbb{R} \), the tempered distribution \( \tau_0(u) \) coincides with a smooth function, \( \tau_0^{\text{loc}}(u) \in C^\infty(\Omega) \), in the sense that for all \( \psi \in \mathcal{D}(\Omega) \), \( \langle \tau_0(u), \psi \rangle = \langle \tau_0^{\text{loc}}(u), \psi \rangle \).

(ii) The function \( \mathbb{R} \to C^\infty(\Omega) : \theta \mapsto \tau_0^{\text{loc}}(u) \) is smooth.

Proof. Abel’s criterion implies that for \( \ell \neq 0 \), the improper integral \( \tau_0^{\text{loc}}(u)(a, \ell) := \frac{1}{2\pi} \int e^{i\xi \ell}(\phi^{-1}_\theta)^* \hat{u}(\xi) \, d\xi \) converges, since, by Lemma 4.3, \( (\phi^{-1}_\theta)^* \hat{u} \in C_0(\mathbb{S}) \). The function \( \tau_0^{\text{loc}}(u) \) is smooth in \((\theta, \ell) \in \mathbb{R} \times \Omega \) where \( \Omega := \{(a, \ell) \mid \ell \neq 0 \} \). Indeed, observing that \( e^{i\xi \ell} = \frac{1}{i\ell} \partial_{\xi} e^{i\xi \ell} \), one obtains, using Lemma 4.3:

\[
\tau_0^{\text{loc}}(u)(a, \ell) = \frac{i}{2\pi \ell} \lim_{r \to \infty} \int_{-r}^{r} e^{i\xi \ell} \partial_{\xi}(\phi^{-1}_\theta)^* \hat{u}(\xi) \, d\xi = \frac{i}{2\pi \ell} \int e^{i\xi \ell} \partial_{\xi}(\phi^{-1}_\theta)^* \hat{u}(\xi) \, d\xi.
\]

Hence, \( \tau_0^{\text{loc}}(u)(a, \ell) = \frac{-1}{2\pi \ell^2} \int e^{i\xi \ell} \partial_{\xi}^2(\phi^{-1}_\theta)^* \hat{u}(\xi) \, d\xi \), which is an existing integral. Now, by Lemma 4.2, one observes that setting \( \hat{u}_\theta := (\phi^{-1}_{1/2})^* \hat{u} \) defines a smooth family of Schwartz functions \( \{ \hat{u}_\theta \} \subset \mathcal{F} \) such that \( (\phi^{-1}_\theta)^* \hat{u} = (\phi^{-1}_{1/2})^* \hat{u}_\theta \). Therefore, one has

\[
\tau_0^{\text{loc}}(u)(a, \ell) = \frac{-1}{2\pi \ell^2} \int e^{i\xi \ell} \partial_{\xi}^2(\phi^{-1}_{1/2})^* \hat{u}_\theta(\xi) \, d\xi,
\]

whose integrand is bounded by the integrable function \( |\partial_{\xi}^2(\phi^{-1}_{1/2})^* \hat{u}_\theta| \) for \( \theta \in [-\epsilon, \epsilon] \).

One proceeds similarly for derivatives. It follows that \( \tau_0^{\text{loc}}(u) \) defines an element of \( C^\infty(\mathbb{R}, C^\infty(\Omega)) \) [Trèves 1967, Theorem 40.1, p. 415]. Of course, \( \langle \tau_0(u), \psi \rangle = \langle \tau_0^{\text{loc}}(u), \psi \rangle \) for all \( \psi \in \mathcal{D}(\Omega) \).

\[\square\]

Proposition 4.5 [Bieliavsky 2002].

(i) The map \( T_\theta := \mathcal{F}^{-1} \circ \phi_\theta^* \circ \mathcal{F} \) establishes a linear isomorphism

\[ T_\theta : \mathcal{E}_\theta \to \mathcal{F}. \]

(ii) Denote by \( \ast^0_\theta \) Weyl’s product on \( \mathcal{F} \). The formula

\[ a \ast^0_\theta b := \tau_0(T_\theta a \ast^0_\theta T_\theta b); \quad a, b \in \mathcal{E}_\theta \]

defines an associative product \( \ast_\theta \) on \( \mathcal{E}_\theta \).

(iii) At the level of smooth compactly supported \( u, v \in C^\infty_c(\mathbb{S}) \) and for all \( x_0 \in \Omega \), one has the oscillatory integral formula

\[ u \ast_\theta v(x_0) = \frac{1}{4\pi^2 \theta^2} \int_{G \times G} \cosh 2(a_1 - a_2) \exp \frac{i}{2\theta} \ell_2 \sinh 2(a_0 - a_1) \, u(x_1) v(x_2) \, dx_1 \, dx_2, \]

where \( x_j = (a_j, \ell_j) \), \( j = 0, 1, 2 \).
(iv) The three-point kernel

\[ K_\theta(x_0, x_1, x_2) := \frac{1}{4\pi^2 \theta^2} \cosh(2(a_1 - a_2)) e^{\frac{\theta}{2} \left( \frac{\epsilon_1}{0,1,2} \sinh(2(a_0 - a_1)) \epsilon_2 \right)} \]

is left-invariant with respect to the diagonal action of \( \mathbb{S} \) on \( G \times G \times G \).

**Remark 4.1.** The formula in (iii) holds at every point \( x_0 \in G \). Indeed, by left-invariance of the kernel, one has for every choice of \( y \in \Omega \): \( u \ast \theta \ast v(x_0) = (L^*_{x_0} \ast u \ast \theta \ast L_{x_0}^{-1} \ast v)(y) \).

We set, for integral \( n \geq 1 \),

\[ \beta(2n + 1) := \frac{1}{2} \frac{(-1)^n}{4^n} \frac{1}{2n + 1} \frac{(2n)!}{(n!)^2} = \frac{(-1)^n}{4^n} \frac{1}{2n + 1} \binom{2n - 1}{n - 1}, \]

and further \( \beta(1) := \frac{1}{2} \) and \( \beta(2n) := 0 \) for \( n \geq 0 \).

**Definition 4.1.** Let \( k \) and \( N \) be positive integers such that \( k \leq N \). We set

\[ B^k_N := \sum_{k_1 + \cdots + k_N = N} \frac{N!}{k_1! k_2! \cdots k_N!} \beta(1)^{k_1} \beta(2)^{k_2} \cdots \beta(N)^{k_N}, \]

with the convention that \( \beta(2n)^0 := 1 \) and \( B^0_0 := 1 \).

**Lemma 4.6.** Let \( u \in C^\infty(\mathbb{S}) \) and \( N \in \mathbb{N}_0 \). Then

\[ \frac{d^N}{d\theta^N} \big|_{\theta = 0} \left( R^*_{\frac{1}{2} \arcsinh, 0} u \right) = \sum_{k=1}^N B^k_N \tilde{H}^k u, \]

where \( R \) denotes right multiplication on \( \mathbb{S} \).

**Proof.** We shall use Faà di Bruno’s formula [1857], which computes the higher order derivatives of composed functions:

\[ (6) \quad \frac{d^n}{d\xi^n} [R(b(\xi))] \]

\[ = \sum_{k = k_1 + \cdots + k_n} \frac{n!}{k_1! k_2! \cdots k_n!} \left( \frac{b^{(1)}(\xi)}{1!} \right)^{k_1} \left( \frac{b^{(2)}(\xi)}{2!} \right)^{k_2} \cdots \left( \frac{b^{(n)}(\xi)}{n!} \right)^{k_n} R^{(k)}(b(\xi)), \]

with \( k := k_1 + \cdots + k_n \) and where \( (k_1, \ldots, k_n) \) runs over all partitions of \( n \) (solutions of the equation \( n = 1k_1 + 2k_2 + \cdots + nk_n \)). We now observe the Taylor expansion

\[ \frac{1}{2} \arcsinh \xi = \frac{\xi}{2} + \sum_{n \geq 1} \frac{(-1)^n}{4^n \ 2n + 1} \left( \frac{2n + 1}{2n - 1} \right)^n \xi^{2n + 1} = \sum_{n \geq 0} \frac{(-1)^n \ 1}{2 \cdot 4^n \ 2n + 1} \left( \frac{2n}{n!} \right)^2 \xi^{2n + 1}. \]

Plugging \( b(\xi) := \frac{1}{2} \arcsinh \xi \) and \( R(a) := R^*_{(\alpha, 0)} u(x) \) into (6) at \( \xi = 0 \) then yields the lemma. \( \square \)
Lemma 4.7. Let $\gamma$ and $\mu$ be maps from $\mathbb{R} \times \mathcal{I}$ to $C^\infty(\mathbb{S})$ such that

(i) for every $u \in \mathcal{I}$ the functions $\gamma(\cdot, u)$ and $\mu(\cdot, u)$ belong to $C^\infty(\mathbb{R}, C^\infty(\mathbb{S}))$;

(ii) for all $\theta$ the partial functions $\gamma(\theta, \cdot)$ and $\mu(\theta, \cdot)$ are $\mathbb{C}$-linear.

Assume the function $\theta \mapsto \phi(\theta, u) := \mu(\theta, \gamma(\theta, u))$ to be an element of $C^\infty(\mathbb{R}, C^\infty(\mathbb{S}))$ as well. Denoting by $\gamma(\theta, u) \sim \sum \theta^m \Gamma_n(u)$, $\mu(\theta, u) \sim \sum \theta^m M_n(u)$, $\phi(\theta, u) \sim \sum \theta^m \Phi_n(u)$ the corresponding Taylor expansions, one has

$$\Phi_n(u) = \sum_{m=n+k} \Phi_n M_n(u).$$

Proof. Let $R^\mu_N(u)$ be the rest at order $N$ in the Taylor expansion of $\mu$ [Bourbaki 1967]. Then,

$$[\theta \mapsto R^\mu_N(\gamma(\theta, u))] \in o(\theta^N).$$

Indeed, $R^\mu_N(u) = \theta^N \rho_N(u)$ where $\rho^N(u)$ is continuous in a neighborhood of $\theta = 0$. Therefore $\theta^{-(N+1)} R^\mu_N(\gamma(\theta, u)) = \sum_{k=N+1} \theta^{k-N-1} \Gamma_k(u) + \theta R^\mu_N(\rho_{N+1}(u))$. The first $N + 2$ terms of this sum tend to 0 with $\theta$. In the last term, one observes that $R^\mu_N(\rho_{N+1}(u))$ is continuous and hence bounded for $\theta \in [-\epsilon, \epsilon]$, which proves the assertion and with it the lemma.

Of course an analogous statement holds for bilinear maps.

Lemma 4.8. For every smooth compactly supported $u, v \in C^\infty_c(\mathbb{S})$, the functions $\theta \mapsto T_\theta(u)$, $u * \theta v$ and $u * \theta v$ belong to $C^\infty(\mathbb{R}, C^\infty(\mathbb{S}))$.

Proof. For every $x_0 \in G$,

$$T_\theta(u)(x_0) = \frac{1}{2\pi} \int e^{i\theta \xi} \delta_{\theta} \hat{u}(a_0, \xi) \, d\xi.$$ 

The integrand $e^{i\theta \xi} \delta_{\theta} \hat{u}(a_0, \xi)$ is smooth in the variables $(a_0, \ell_0, \theta)$ as well as every of its partial derivatives. Moreover, each of these functions is bounded by an integrable function since $\delta_{\theta} \hat{u}$ is Schwartz. Therefore the function $(x_0, \theta) \mapsto T_\theta(u)(x_0)$ is smooth.

The argument is similar for the other functions, simpler in fact since the integrand is in both cases compactly supported.

Theorem 4.9. For every smooth compactly supported $u, v \in C^\infty_c(\mathbb{S})$ and for every $x \in G$, the function $[\mathbb{R} \to \mathbb{C} : \theta \mapsto u * \theta v(x)]$ is smooth. Its Taylor series at 0 defines
an associative formal star product $\star_v$ on $C^\infty(\mathbb{S})[v]$ whose explicit expression is

$$u \star_v v := u + \sum_{n=1}^{\infty} \sum_{n=1}^{\infty} \frac{v^n}{n!} (-2i)^n \sum_{k=0}^{n} (-1)^{k+1} \binom{n}{k}$$

$$\times \left( \left( \sum_{m=0}^{k} B^m_k \tilde{H}^m \tilde{E}^{n-k} \right) \left( \sum_{m=0}^{n-k} B^m_{n-k} \tilde{H}^m \tilde{E}^k \right) - 2 \left( \sum_{r=0}^{k} \binom{k}{r} (k-r) \beta(k-r) \sum_{p=0}^{r} B^p_r \tilde{H}^p \tilde{E}^{n-k} \right) \times \left( \sum_{s=0}^{n-k} \binom{n-k}{s} (n-k-s) \beta(n-k-s) \sum_{q=0}^{s} B^q_s \tilde{H}^q \tilde{E}^k \right) \right).$$

**Proof.** Using left invariance one gets

$$u \star_v v(x_0) = \int K_\theta(0, x_1, x_2) R_{x_1}^* u \big|_{x_0} R_{x_2}^* v \big|_{x_0} d x_1 d x_2.$$ 

Making the change of variables $\xi := \sinh(2\alpha), \eta := \frac{1}{2\theta} \ell_i$ yields

$$u \star_v v(x_0) = \frac{1}{4\pi^2} \int \left( 1 - \gamma(\xi_1) \gamma(\xi_2) \right) e^{-i(\xi_1 \eta_2 - \xi_2 \eta_1)} \times r_{(\xi_1, 2\theta \eta_1)}^* u \big|_{x_0} r_{(\xi_2, 2\theta \eta_2)}^* v \big|_{x_0} d \xi_1 d \xi_2 d \eta_1 d \eta_2,$$

where $\gamma(\xi) := \xi / \sqrt{1 + \xi^2}$ and $r_{(\xi, 2\theta \eta)}^* u \big|_x := u(a + \frac{1}{2} \arcsinh \xi, e^{-\arcsinh \xi} \ell + 2\theta \eta)$.

Now observe the asymptotic expansion ($\theta_0 \neq 0$)

$$u \star_v v(x_0) = \sum_{n=0}^{\infty} \frac{(\theta - \theta_0)^n}{n!} \frac{1}{4\pi^2} \int \left( 1 - \gamma(\xi_1) \gamma(\xi_2) \right) e^{-i(\xi_1 \eta_2 - \xi_2 \eta_1)} \times \left[ r_{(\xi_1, 2\theta \eta_1)}^* u \big|_{x_0} r_{(\xi_2, 2\theta \eta_2)}^* v \big|_{x_0} \right] d \xi_1 d \xi_2 d \eta_1 d \eta_2 + o((\theta - \theta_0)^N).$$

One has

$$\frac{d^n}{d\theta^n} \left[ r_{(\xi_1, 2\theta \eta_1)}^* u \big|_{x_0} r_{(\xi_2, 2\theta \eta_2)}^* v \big|_{x_0} \right] \bigg|_{\theta = \theta_0}$$

$$= \sum_{k=0}^{n} \binom{n}{k} \frac{d^k}{d\theta^k} \left[ r_{(\xi_1, 2\theta \eta_1)}^* u \big|_{x_0} \right] \bigg|_{\theta = \theta_0} \frac{d^{n-k}}{d\theta^{n-k}} \left[ r_{(\xi_2, 2\theta \eta_2)}^* v \big|_{x_0} \right] \bigg|_{\theta = \theta_0}$$

$$= 2^n \sum_{k=0}^{n} \binom{n}{k} \eta_1^k \eta_2^{n-k} r_{(\xi_1, 2\theta \eta_1)}^* \left( \tilde{E}^k \right) \big|_{x_0} r_{(\xi_2, 2\theta \eta_2)}^* \left( \tilde{E}^{n-k} \right) \big|_{x_0},$$
where \( \exp(aH) := (a, 0) \) and \( \exp(\ell E) := (0, \ell) \). Therefore, an integration by parts yields (for \( n \geq 1 \))

\[
\int (1 - \gamma(\xi_1)\gamma(\xi_2)) e^{-i(\xi_2-\xi_1)} \frac{d^n}{d\eta^n} [n_{(\xi_1, 2\theta \eta_1)}^* r_{(\xi_2, 2\theta \eta_2)}^* v]_{x_0} |_{\theta = 0} d\xi_1 d\eta_1 d\eta_2
\]

is seen to be

\[
4\pi^2 (-2i)^n \sum_{k=0}^n (-1)^{k+1} \binom{n}{k} \gamma(\xi_2) r_{(\xi_2, 2\theta \eta_2)}^* (\tilde{E}^{n-k} v) |_{x_0}
\]

By observing that \( \lim_{\eta \to 0} \partial_{\xi} r_{(\xi, 0)}^* u |_{x} = \partial_{\xi} r_{(\xi, 0)}^* u |_{x} \), one gets

\[
\lim_{\theta \to 0} \int \gamma(\xi_1)\gamma(\xi_2) e^{-i(\xi_2-\xi_1)} \frac{d^n}{d\eta^n} [n_{(\xi_1, 2\theta \eta_1)}^* r_{(\xi_2, 2\theta \eta_2)}^* v]_{x_0} d\xi_1 d\eta_1 d\eta_2
\]

which, by using Lemma 4.6 and the expansion

\[
\gamma(\xi) = \xi + \sum_{m \geq 1} \frac{(-1)^m m + 1}{4m} \binom{2m}{m + 1} \xi^{2m + 1} = \sum_{m \geq 0} \frac{(-1)^m (2m)!}{4m (m!)^2} \xi^{2m + 1},
\]

is seen to be

\[
4\pi^2 (-2i)^n \sum_{k=0}^n (-1)^{k+1} \binom{n}{k} \left( \sum_{r=0}^k \binom{k}{r} \gamma^{(k-r)} (0) \sum_{p=0}^{r} B_p^r (\tilde{H}^p \tilde{E}^{n-k} v) |_{x_0} \right)
\]

This leads to the announced asymptotic expansion where \( v \) is the formal parameter corresponding to \( \theta \).

Regarding associativity, one first observes that for all \( x_0 \in \Omega \), one has the coincidence: \( u \ast_0 v(x_0) = r_0^{loc}(T_0 u \ast_0 T_0 v)(x_0) \). Note that, independently, the left hand side is smooth in \( \theta \) by virtue of Lemma 4.8, while the right hand side is smooth as well by virtue of Lemmas 4.4, 4.7 and 4.8. Lemma 4.7 therefore identifies the above-computed Taylor coefficients with the cochains of the star product obtained by intertwining Moyal’s star product with a formal equivalence. \( \square \)
Remark 4.2. Our construction differs radically from the (formal) universal deformation formula of Giaquinto and Zhang [1998]: It was shown in [Bieliavsky et al. 2005b] that the (maximal) invariance diffeomorphism group of Giaquinto–Zhang star product on $ax + b$ is isomorphic to $Sp(1, \mathbb{R}) \times \mathbb{R}^2$, while in our case the corresponding maximal invariance group is the automorphism group of the underlying symplectic symmetric space — in the two-dimensional case, the solvable group $SO(1, 1) \times \mathbb{R}^2$.

5. Hopf structure

The formal star product $\star_v$ given by formula (7) can be described as an example of the following construction [Bieliavsky et al. 2005a].

Definition 5.1 [Bonneau and Sternheimer 2005]. Let $B$ be a cocommutative bialgebra and $C$ a $B$-bimodule algebra. The $L$-$R$-smash product $C \bowtie B$ is the algebra constructed on the vector space $C \otimes B$ where the multiplication is defined by

$$
(f \otimes a) \star (g \otimes b) = \sum_{(a)(b)} (f \leftarrow b_{(1)}) (a_{(1)} \rightarrow g) \otimes a_{(2)} b_{(2)}
$$

for $f, g \in C$ and $a, b \in B$.

On the first hand, the above construction provides deformation quantizations of $\mathcal{C}^\infty(G) \otimes \text{Pol}(\mathfrak{g}) \subset \mathcal{C}^\infty(T^*G)$, where $\mathfrak{g} = \text{lie}(G)$, as explained in [Bieliavsky et al. 2005a]. For $G = \mathbb{R}^n$ it reproduces the Moyal star product. On the other hand we have seen that, for $u, v \in \mathcal{C}^\infty(\mathbb{R}^2)$, $u \star_v v = T_v^{-1}(T_v u \star^0_v T_v v)$ where $\star^0_v$ denotes the Moyal star product on $\mathcal{C}^\infty(\mathbb{R}^2)$. Restricting to $\mathcal{C}^\infty(\mathbb{R}) \otimes \text{Pol}(\mathbb{R}) \subset \mathcal{C}^\infty(\mathbb{R}^2)$, the formal equivalence $T_v$ is viewed as $id \otimes S_v$.

Proposition 5.1. The formal star product $\star_v$ given by (7) coincides with the product underlyg the $L$-$R$-smash product $\mathcal{C}^\infty(\mathbb{R})[v] \bowtie \text{Pol}(\mathbb{R})$ with $x \rightarrow f = f \leftarrow x = \frac{1}{2} S_v^{-1}(\frac{d}{dx}(S_v f))$ and we consider $\text{Pol}(\mathbb{R})$ endowed with its usual Hopf structure.

Proof. See [Bieliavsky et al. 2005a]: the main fact is that $\mathcal{C}^\infty(\mathbb{R})[v] \bowtie \text{Pol}(\mathbb{R})$ carries the Moyal product for $x \rightarrow f = f \leftarrow x = \frac{1}{2} \frac{d}{dx}(f)$.

F. Panaite and F. Van Oystaeyen [2005] have shown that if $B$ be a cocommutative Hopf algebra and $C$ a $B$-module algebra, the map $\Phi : C \bowtie B \rightarrow C \bowtie B$ with $\Phi(f \bowtie a) = \sum_{(a)} f \leftarrow a_{(1)} \# a_{(2)}$ is an isomorphism of algebras, the multiplication on the smash product $C \bowtie B$ being given by $(f \otimes a) \star (g \otimes b) = \sum_{(a)} (a_{(1)} \bullet g) \otimes a_{(2)} b$ with $a \bullet g = a_{(1)} \rightarrow g \leftarrow J(a_{(2)})$, $J$ the antipode of $B$. One therefore has

Proposition 5.2. The star product $\star_v$ can also be seen as a classical smash product in the sense of Sweedler [1968].

At last, it is shown in [Bieliavsky et al. 2005a] that, under conditions, if $C$ is a Hopf algebra then so is $C \bowtie B$. Seing $\mathcal{C}^\infty(\mathbb{R})[v]$ as $\mathcal{C}^\infty(\mathbb{R}) \otimes \mathbb{R}[v]$ and considering
the natural Hopf structure on the tensor product of the Hopf algebras \( \mathcal{C}^\infty(\mathbb{R}) \) and \( \mathbb{R}[\nu] \) we show in [Bieliavsky et al. 2005a] that the needed conditions are fulfilled. Hence:

**Theorem 5.3.** There exists a coproduct \( \Delta_v \), a counit \( \epsilon_v \) and an antipode \( J_v \), compatible with \( \star_v \), such that \( \left( \mathcal{C}^\infty(\mathbb{R})[\nu] \otimes \mathfrak{pol}(\mathbb{R}), \star_v, \Delta_v, \epsilon_v, J_v \right) \) is a Hopf algebra.

Replacing \( \mathbb{R} \) by \( \mathbb{R}^n \) these structural results still hold on the asymptotic expansion of the product described in formula (4).

**Acknowledgment**

We thank the referee for suggesting improvements to the original manuscript.

**References**


Received March 25, 2005.

PIERRE BIELIAVSKY
DÉPARTEMENT DE MAThÉMATIQUE
UNIVERSITÉ CATHOLIQUE DE LOUVAIN
CHEMIN DU CYCLOTRON, 2
1348 LOUVAIN-LA-NEUVE
BELGIUM
bieliavsky@math.ucl.ac.be

PHILIPPE BONNEAU
LABORATOIRE DE MAThÉMATIQUE ET APPLICATIONS DE METZ
UMR CNRS 7122, UNIVERSITé DE METZ
ILE DU SAULCY
F-57045 METZ CEDEX 1
FRANCE
bonneau@univ-metz.fr

YOSHIAKI MAEDA
DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE AND TECHNOLOGY
KEIO UNIVERSITY
3-14-1, HIYOSHI, KOHOKU-KU
YOKOHAMA, 223-8522
JAPAN
maeda@math.keio.ac.jp
Oscillation and nonoscillation properties of second order Sturm–Liouville dynamic equations on time scales—for example, second order self-adjoint differential equations and second order Sturm–Liouville difference equations—have attracted much interest. Here we consider a given homogeneous equation and a corresponding equation with forcing term. We give new conditions implying that the latter equation inherits the oscillatory behavior of the homogeneous equation. We also give new conditions that introduce oscillation of the inhomogeneous equation while the homogeneous equation is nonoscillatory. Finally, we explain a gap in a result given in the literature for the continuous and the discrete case. A more useful result is presented, improving the theory even for the corresponding continuous and discrete cases. Examples illustrating the theoretical results are supplied.

1. Introduction

The theory of dynamic equations on time scales continues to be a rapidly growing area of research. Behind the main motivation for the subject lies the key concept that dynamic equations on time scales represent a way of unifying and extending continuous and discrete analysis. In this paper, we consider the second order linear dynamic equation

\[(c(t) x^\Delta)^\Delta + q(t) x^\sigma = 0\]

(1)

together with an inhomogeneous equation of the form

\[(c(t) u^\Delta)^\Delta + q(t) u^\sigma = f(t).\]

(2)

Equations (1) and (2) are so-called dynamic equations on a time scale \(\mathbb{T}\). Throughout this paper we assume that \(c\), \(q\), and \(f\) are rd-continuous real-valued functions defined on the time scale \(\mathbb{T}\) such that \(c(t) \neq 0\) for all \(t \in \mathbb{T}\) and \(f\) is not eventually

**MSC2000:** primary 34C10, 34K11, 39A11; secondary 39A10, 39A12.

**Keywords:** dynamic equation, generalized zero, oscillation, nonoscillation, inhomogeneous equation, time scale.

Supported by the Australian Research Council’s Discovery Projects (DP0450752).
identically equal to zero. No further assumptions on the sign of these functions are imposed. Since we are interested in oscillatory behavior of solutions of (1) and (2), we assume that the time scale $\mathbb{T}$ is unbounded above. The setup of this paper is as follows. In Section 2, we give some preliminaries concerning the time scales calculus. In Section 3, we introduce the Komkov transformation and present some basic results about equations (1) and (2). In Section 4, given that (1) is nonoscillatory, we offer criteria that introduce oscillation in (2) and also criteria that preserve nonoscillation in (2). Finally, in Section 5, we explain a gap in a result given in the continuous case by Rankin [1979, Theorem 1] and in the discrete case by Grace and El-Morshedy [1997, Theorem 2.1]. A more useful result is presented, hence improving the theory even for the corresponding continuous and discrete cases. Throughout, relevant examples illustrating the theoretical results are supplied.

2. The time scales calculus

In this section we present some definitions and elementary results connected to the time scales calculus. For further study we refer the reader to [Bohner and Peterson 2001; 2003]. A time scale $\mathbb{T}$ is an arbitrary nonempty closed subset of the real numbers $\mathbb{R}$. On $\mathbb{T}$ we define the forward and backward jump operators by

$$\sigma(t) := \inf \{ s \in \mathbb{T} : s > t \} \quad \text{and} \quad \rho(t) := \sup \{ s \in \mathbb{T} : s < t \} \quad \text{for} \quad t \in \mathbb{T}.$$ 

A point $t \in \mathbb{T}$ with $t > \inf \mathbb{T}$ is said to be left-dense if $\rho(t) = t$ and right-dense if $\sigma(t) = t$, left-scattered if $\rho(t) < t$ and right-scattered if $\sigma(t) > t$. Next, the graininess function $\mu$ is defined by $\mu(t) := \sigma(t) - t$ for $t \in \mathbb{T}$. For a function $f : \mathbb{T} \to \mathbb{R}$ the (delta) derivative $f^\Delta(t)$ at $t \in \mathbb{T}$ is defined to be the number (provided it exists) with the property such that for every $\varepsilon > 0$ there exists a neighborhood $U$ of $t$ with

$$|f(\sigma(t)) - f(s) - f^\Delta(t)(\sigma(t) - s)| \leq \varepsilon |\sigma(t) - s| \quad \text{for all} \quad s \in U.$$

A useful formula is

$$(3) \quad f^{\sigma} = f + \mu f^\Delta, \quad \text{where} \quad f^{\sigma} := f \circ \sigma.$$ 

We will use the product rule and the quotient rule for the derivative of the product $fg$ and the quotient $f/g$ (if $gg^{\sigma} \neq 0$) of two differentiable functions $f$ and $g$

$$(4) \quad (fg)^\Delta = f^\Delta g + f^{\sigma} g^\Delta = f g^\Delta + f^\Delta g^{\sigma} \quad \text{and} \quad \left( \frac{f}{g} \right)^\Delta = \frac{f^\Delta g - g^\Delta f}{gg^{\sigma}}.$$ 

For $a, b \in \mathbb{T}$ and a function $f : \mathbb{T} \to \mathbb{R}$, the Cauchy integral of $f$ is defined by

$$(5) \quad \int_a^b f(t) \Delta t = F(b) - F(a),$$ 

where $F$ is the antiderivative of $f$. A more useful result is presented, hence improving the theory even for the continuous and discrete cases. Throughout, relevant examples illustrating the theoretical results are supplied.
where $F$ is an antiderivative of $f$, i.e., $F^\Delta = f$ holds. The function $f : \mathbb{T} \to \mathbb{R}$ is called rd-continuous if it is continuous in right-dense points and if the left-sided limits exist in left-dense points. Hilger’s main existence theorem [Bohner and Peterson 2001, Theorem 1.74] says that rd-continuous functions possess antiderivatives. If $p : \mathbb{T} \to \mathbb{R}$ is rd-continuous and regressive (i.e., $1 + \mu(t)p(t) \neq 0$ for all $t \in \mathbb{T}$), then another existence theorem says that the initial value problem $y^\Delta = p(t)y, y(t_0) = 1$ (where $t_0 \in \mathbb{T}$) possesses a unique solution $e_p(\cdot, t_0)$.

**Example 1.** Note that in the case $\mathbb{T} = \mathbb{R}$ we have

$$
\sigma(t) = t, \quad \mu(t) \equiv 0, \quad f^\Delta(t) = f'(t),
$$

and in the case $\mathbb{T} = \mathbb{Z}$ we have

$$
\sigma(t) = t + 1, \quad \mu(t) \equiv 1, \quad f^\Delta(t) = \Delta f(t) = f(t + 1) - f(t).
$$

Another important time scale is $\mathbb{T} = q^{\mathbb{N}_0} := \{q^k : k \in \mathbb{N}_0\}$ with $q > 1$, for which

$$
\sigma(t) = qt, \quad \mu(t) = (q - 1)t, \quad f^\Delta(t) = \frac{f(qt) - f(t)}{(q - 1)t},
$$

and this time scale gives rise to so-called $q$-difference equations.

3. Generalized zeros and the Komkov transformation

We say that a solution $x$ of (1) (or (2)) has a generalized zero in $[t, \sigma(t)]$ if

$$
c(t)x(t)x(\sigma(t)) \leq 0.
$$

Next, $x$ is called oscillatory provided $[T, \infty)$ contains infinitely many zeros for each $T \in \mathbb{T}$. Otherwise we say that $x$ is nonoscillatory. The equation (1) (or (2)) is called oscillatory if all solutions of (1) (or (2)) are oscillatory. Otherwise we say that (1) (or (2)) is nonoscillatory. It is a well-known fact that (1) is oscillatory if and only if it has an oscillatory solution. The proof is easy: Suppose $x$ is a nonoscillatory solution of (1), i.e., $cxx^\sigma > 0$ on $[T, \infty)_\mathbb{T}$ for some $T > 0$. Let $\tilde{x}$ be any solution of (1) such that $x$ and $\tilde{x}$ are linearly independent. Then $(\tilde{x}/x)^\Delta = W(x, \tilde{x})/(cxx^\sigma)$ by the quotient rule (4), where $W(x, \tilde{x}) := c(\tilde{x}^\Delta x - x^\Delta \tilde{x})$, the Wronskian, is actually equal to a nonzero constant (use the product rule (4) to verify this). Hence $\tilde{x}/x$ is eventually strictly monotone, and therefore it is eventually of one sign. Thus $(c\tilde{x}\tilde{x}^\sigma)/(cxx^\sigma) = (\tilde{x}/x)(\tilde{x}^\sigma/x^\sigma)$ is eventually positive, and hence $c\tilde{x}\tilde{x}^\sigma > 0$ eventually, meaning that $\tilde{x}$ is nonoscillatory as well.

In contrast to (1), it is not true that (2) is oscillatory if and only if it has an oscillatory solution. We supply the following examples.

**Example 2.** Suppose $x$ solves (1) such that $|x(t)| \leq 1$ for all $t \in \mathbb{T}$ and such that for all $t \in \mathbb{T}$ there exist $t_1, t_2 \geq t$ with $x(t_1) = 1$ and $x(t_2) = -1$. Then $u_1 := 1 + x/2$
is a nonoscillatory solution of (2) with \( f = q \) while \( u_2 := 1 + 2x \) is an oscillatory solution.

**Example 3.** Consider the second order linear dynamic equation

\[
(\mu(t)u^\Delta)^\Delta + \frac{4}{\mu(t)}u^\sigma = \frac{4}{\mu(t)}
\]

on an isolated time scale (i.e., each point is left-scattered and right-scattered). A solution of the corresponding homogeneous equation is \( x = e^{-2/\mu(\cdot, t_0)} \), where \( t_0 \in \mathbb{T} \). Since \( x^\sigma = -x \), this solution is oscillatory. Hence the corresponding homogeneous equation is oscillatory. However, \( u_1 := 1 + x/2 \) is a nonoscillatory solution of (6) while \( u_2 := 1 + 2x \) is an oscillatory solution of (6). Hence (6) possesses both oscillatory and nonoscillatory solutions.

The transformation \( u = xy \), where \( x \) solves (1) and \( u \) solves (2), was studied by Komkov [1972] and has been successfully applied, for example, in [Grace and El-Morshedy 1997; Patula 1979; Rankin 1979]. Our results given in this paper mainly rely on the following easy but useful identity. We abbreviate the operator \((cx^\Delta)^\Delta + qx^\sigma\) by \( Lx \).

**Lemma 1.** If \( u = xy \), then

\[
W(x, u) = cxx^\sigma y^\Delta
\]

and

\[
[W(x, u)]^\Delta = x^\sigma Lx - u^\sigma Lx.
\]

In particular, if \( x \) solves (1) and \( u \) solves (2), then

\[
[W(x, u)]^\Delta = (cxx^\sigma y^\Delta)^\Delta = fx^\sigma,
\]

and if in addition \( x(t) \neq 0 \) for all \( t \geq T \), then

\[
y(t) = y(T) + c(T)x(T)x(\sigma(T))y^\Delta(T) \int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))} \\
+ \int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau)) \Delta \tau \Delta s \quad \text{for all} \quad t \geq T.
\]

**Proof.** We apply the product rule (4) to \( u = xy \) to find \( u^\Delta = x^\Delta y + x^\sigma y^\Delta \) and

\[
 cxx^\sigma y^\Delta = cxx^\Delta y = cx^\Delta x - cxx^\Delta u = W(x, u).
\]

Then, using the product rule again, we find

\[
(cxx^\sigma y^\Delta)^\Delta = (cx^\Delta)^\Delta x^\sigma + cx^\Delta x^\Delta - (cx^\Delta)^\Delta u^\sigma - cx^\Delta u^\Delta
\]

\[
= (Lu - q x^\sigma) x^\sigma - (Lx - q x^\sigma) u^\sigma
\]

\[
= x^\sigma Lu - u^\sigma Lx.
\]
Hence, if $Lx = 0$ and $Lu = f$, (8) follows. By using the definition (5) of the integral, we conclude that (9) holds.

Concluding this section, we use Lemma 1 to derive the following result. For the continuous version see [Rankin 1979, Equation (3')] and for the discrete version see [Patula 1979, Theorem 6] and [Grace and El-Morshedy 1997, Lemma 2.1].

**Theorem 1.** Suppose $x$ solves (1) and $u$ solves (2). If $W(x, u)$ is eventually of one sign (either positive or negative), then $x$ oscillates if and only if $u$ oscillates.

**Proof.** Using (7), we know that $W(x, u)$ is eventually of one sign, where $u = xy$. First suppose $x$ is not oscillatory, i.e., $cxx^\sigma > 0$ eventually. Hence $y^\Delta$ is eventually of one sign. Therefore $y$ is eventually of one sign. Hence $0 < yy^\sigma = \frac{u u^\sigma}{x x^\sigma} = \frac{cuu^\sigma}{cxx^\sigma}$ eventually, so eventually $cuu^\sigma > 0$, i.e., $u$ is not oscillatory. Similarly (by considering the transformation $x = u\tilde{y}$) we may show that if $u$ is not oscillatory, then $x$ is not oscillatory either.

**Corollary 1.** Suppose $c(t) > 0$ for all $t \in \mathbb{T}$. If (1) is nonoscillatory and $f$ is eventually of one sign, then (2) is nonoscillatory.

**Proof.** Let $x$ be any (nonoscillatory) solution of (1) so $x$ is eventually of one sign. Suppose $u$ is any solution of (2) and let $y = u/x$. By (8), $[W(x, u)]^\Delta$ is eventually of one sign, and hence $W(x, u)$ is eventually of one sign. Thus $u$ is nonoscillatory according to Theorem 1.

**Corollary 2.** Suppose (1) is oscillatory (nonoscillatory). If there exists a solution $x$ of (1) such that

$$\int_T^\infty f(t)x(\sigma(t))\Delta t = \infty \quad \text{or} \quad \int_T^\infty f(t)x(\sigma(t))\Delta t = -\infty,$$

then (2) is oscillatory (nonoscillatory).

**Proof.** Suppose $u$ is a solution of (2) and define $y$ by $u = xy$. By (8),

$$W(x, u)(t) = W(x, u)(T) + \int_T^t f(s)x(\sigma(s))\Delta s.$$

Hence $W(x, u)$ is eventually of one sign, and the claim follows with Theorem 1.

**Example 4.** Consider the Fibonacci difference equation

$$x(t+2) = x(t+1) + x(t), \quad \text{i.e.,} \quad \Delta((-1)^t x(t)) + (-1)^t x(t+1) = 0, \quad t \in \mathbb{N}.$$  

If $a = (1 + \sqrt{5})/2$, then $x(t) = a^t$ is a solution of this equation. Since

$$c(t)x(t)x(t+1) = (-a)(-a^2)^t,$$
the equation is oscillatory. Now, since \( \sum_{r=0}^{\infty} a^{r+1} = \infty \), Corollary 2 implies that \( x(t+2) = x(t+1) + x(t) + (-1)^t \), i.e., \( \Delta((-1)^{t+1} \Delta x(t)) + (-1)^{t+1} x(t+1) = 1 \), is also oscillatory.

4. Oscillation and nonoscillation criteria

The next theorem generalizes a result due to Rankin for \( \mathbb{T} = \mathbb{R} \) [Rankin 1979, Theorem 2] and a result due to Grace and El-Morshedy for \( \mathbb{T} = \mathbb{Z} \) [Grace and El-Morshedy 1997, Theorem 2.2].

**Theorem 2.** Suppose \( x \) is an eventually nonoscillatory solution of (1). If for some sufficiently large \( T \in \mathbb{T} \),

(10) \[ \liminf_{t \to \infty} \int_{T}^{t} \frac{\Delta t}{c(t)x(t)x(\sigma(t))} < \infty, \]

(11) \[ \limsup_{t \to \infty} \int_{T}^{t} \frac{1}{c(s)x(s)x(\sigma(s))} \int_{s}^{T} f(\tau)x(\sigma(\tau)) \Delta \tau \Delta s = -\infty, \]

and

(12) \[ \limsup_{t \to \infty} \int_{T}^{t} \frac{1}{c(s)x(s)x(\sigma(s))} \int_{s}^{T} f(\tau)x(\sigma(\tau)) \Delta \tau \Delta s = \infty, \]

then (2) is oscillatory.

**Proof.** Suppose \( u \) is an eventually nonoscillating solution of (2) such that \( y = u/x \) is eventually of one sign (note that \( yy^{\sigma} = (c uu^{\sigma})/(c xx^{\sigma}) > 0 \)). But (9) together with (10), (11), and (12) ensures that

\[ \liminf_{t \to \infty} y(t) = -\infty \quad \text{and} \quad \limsup_{t \to \infty} y(t) = \infty. \]

This is a contradiction, and therefore there cannot exist an eventually nonoscillating solution of (2). Thus (2) is oscillatory. \( \square \)

**Example 5.** Let \( q > 1 \) and consider the \( q \)-difference equation (see Example 1)

(13) \[ u^{\Delta \Delta} = (-1)^{\log_q t}, \quad t \in q^{n_0} := \left\{ q^k : k \in \mathbb{N}_0 \right\}. \]

One solution of the corresponding homogeneous equations is \( x(t) = t \), so the homogeneous equation is nonoscillatory and (10) is satisfied since

\[ \int_{1}^{t} \frac{\Delta s}{c(s)x(s)x(\sigma(s))} = 1 - \frac{1}{t} \to 1 \quad \text{as} \quad t \to \infty. \]

Some calculation now shows that

\[ \int_{1}^{t} \frac{1}{q s^2} \int_{s}^{T} q \tau (-1)^{\log_q \tau} \Delta \tau \Delta s = \frac{(q - 1)^2}{q^3 + q^2 + q + 1} t(1)^{\log_q t}, \]
so (11) and (12) are satisfied. Hence, by Theorem 2, (13) is oscillatory.

Now we present an improvement of Theorem 1. For $\mathbb{T} = \mathbb{Z}$, see [Grace and El-Morshedy 1997, Theorem 2.3].

**Theorem 3.** Suppose $x$ solves (1) such that $x(t) \neq 0$ for all $t \geq T$ and

\[
\int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))} \quad \text{is bounded above or below.}
\]

If

\[
\int_T^\infty \frac{1}{c(t)x(t)x(\sigma(t))} \int_T^t f(s)x(\sigma(s)) \Delta s \Delta t = \infty,
\]

then (1) and (2) either are both oscillatory or both nonoscillatory.

**Proof.** Let $u$ be a solution of (2). By Theorem 1, we may assume that $W(x, u)$ is oscillating. First suppose that the integral in (14) is bounded below. Let $T \in \mathbb{T}$ and $D > 0$ be such that

\[
W(x, u)(T) \geq 0 \quad \text{and} \quad \int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))} \geq -D \quad \text{for all} \quad t \geq T.
\]

Then, by (9),

\[
y(t) \geq y(T) - DW(x, u)(T) + \int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau)) \Delta \tau \Delta s,
\]

so $y(t) \to \infty$ as $t \to \infty$ by (15). Hence $y > 0$ eventually and $cux^\sigma$ has eventually the same sign as $cx\sigma$. If, however, the integral in (14) is bounded above, then we pick $T \in \mathbb{T}$ and $E > 0$ such that

\[
W(x, u)(T) \leq 0 \quad \text{and} \quad \int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))} \leq E \quad \text{for all} \quad t \geq T.
\]

In this case the conclusion now follows as in the previous case. \qed

The last result in this section is a nonoscillation criterion. We refer to [Grace and El-Morshedy 1997, Theorem 3.1] for $\mathbb{T} = \mathbb{Z}$. The following auxiliary result is needed.

**Lemma 2.** Suppose (1) is nonoscillatory. Then there exists a solution $x$ of (1) satisfying (10).

**Proof.** Let $x$ be any (nonoscillatory) solution of (1). If $x$ satisfies (10), then we are done. If not, then $\int_T^\infty \frac{\Delta s}{c(s)x(s)x(\sigma(s))} = \infty$. Let $\tilde{x}$ be any solution of (1) such that $x$ and $\tilde{x}$ are linearly independent, i.e., $W(x, \tilde{x}) \equiv -k \neq 0$. Then
\((\tilde{x}/x)^\Delta = k/(cxx^\sigma)\), so \((\tilde{x}/x)(t) \to \pm\infty\) as \(t \to \infty\) and hence \((x/\tilde{x})(t) \to 0\) as \(t \to \infty\). Thus \((x/\tilde{x})^\Delta = -k/(c\tilde{x}\tilde{x}^\sigma)\) and therefore
\[
k \int_T^t \frac{\Delta s}{c(s)\tilde{x}(s)\tilde{x}(\sigma(s))} = \frac{x(T)}{\tilde{x}(T)} - \frac{x(t)}{\tilde{x}(t)} \to \frac{x(T)}{\tilde{x}(T)} \quad \text{as} \quad t \to \infty.
\]
Hence \(\tilde{x}\) solves (1) and satisfies (10). \(\square\)

Below we use for \(\alpha \in \mathbb{R}\) the notation \(\alpha^+ = \max\{0, \alpha\}\) and \(\alpha^- = \min\{0, \alpha\}\).

**Theorem 4.** Suppose \(c(t) > 0\) and \(q(t) > 0\) for all \(t \in \mathbb{T}\). Suppose that all solutions of (1) are nonoscillatory and bounded. If
\[
\int_T^t \frac{1}{c(s)} \int_T^s f^+(\tau) \Delta \tau \Delta s = \infty \quad \text{(16)}
\]
and
\[
\int_T^\infty -f^-(\tau) \Delta \tau > -\infty \quad \text{(17)}
\]
then (2) is nonoscillatory.

**Proof.** Let \(\lambda > 0\) be such that the integral in (17) is bounded below by \(-\lambda\). By Lemma 2, there exists a solution \(x\) of (1) satisfying (10). Since all solutions of (1) are bounded, there exists \(M > 0\) such that \(|x(t)| < M\) for all \(t \in \mathbb{T}\). We will show that (15) is satisfied. Then Theorem 3 is employed to complete the proof.

First, putting \(z = cx^\Delta\), we see that \(z^\Delta = -qx^\sigma\) is eventually of one sign. Thus \(z\) is eventually of one sign. Hence \(x^\Delta\) is eventually of one sign. Therefore \(x\) is eventually increasing or eventually decreasing. Since \(x\) is nonoscillatory, it is either eventually positive or eventually negative. So there are the following four possibilities: (i) \(x\) is eventually positive and increasing; (ii) \(x\) is eventually positive and decreasing; (iii) \(x\) is eventually negative and increasing; (iv) \(x\) is eventually negative and decreasing. If (iii) or (iv) holds, then we may replace \(x\) by \(-x\), which is also a solution of (1) that satisfies (10) and (i) or (ii). Thus it is sufficient to discuss the cases (i) and (ii). If (i) holds, then
\[
\int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau)) \Delta \tau \Delta s \geq \frac{x(\sigma(T))}{M^2} \int_T^t \frac{1}{c(s)} \int_T^s f^+(\tau) \Delta \tau \Delta s - \lambda M \int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))} \to \infty \quad \text{as} \quad t \to \infty
\]
(use \( f = f^+ + f^- \)), while if (ii) holds, then
\[
\int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau))d\tau ds \\
\geq \int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f^+(\tau)x(\sigma(s))d\tau ds - \lambda \int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))} \\
\geq \frac{1}{M} \int_T^t \frac{1}{c(s)} \int_T^s f^+(\tau)d\tau ds - \lambda \int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))}
\]

\( \text{(16)} \quad \lim_{t \to \infty} \int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))} \to \infty \). Hence (15) holds in either case and the proof is complete. \( \square \)

5. Remarks on the results of Rankin, Grace and El-Morshedy

The continuous version of the following result was proved by Rankin [1979, Theorem 1], while its discrete version was given by Grace and El-Morshedy in [1997, Theorem 2.1].

**Theorem 5.** Suppose \( x \) is an eventually nonoscillatory solution of (1). If for sufficiently large \( T \in \mathbb{T} \) and some \( M > 0 \),

\[
\text{(18)} \quad \lim_{t \to \infty} \int_T^t \frac{\Delta t}{c(t)x(t)x(\sigma(t))} = \infty,
\]

\[
\text{(19)} \quad \liminf_{t \to \infty} \int_T^t f(s)x(\sigma(s))ds = -\infty, \quad \limsup_{t \to \infty} \int_T^t f(s)x(\sigma(s))ds = \infty,
\]

and

\[
\text{(20)} \quad \left| \int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau))d\tau ds \right| \leq M \int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))}
\]

for all \( t \geq T \), then (2) is oscillatory.

Below we show that the assumptions of Theorem 5 are never satisfied. This means that, although Theorem 5 is true, the result is not meaningful.

**Theorem 6.** The assumptions of Theorem 5 are never satisfied.

**Proof.** Note first that (2) has a solution. Let \( u \) be any solution of (2) and define \( y \) by \( u = xy \). Let \( T \in \mathbb{T} \) such that \( c(t)x(t)x(\sigma(t)) \geq 0 \) for all \( t \geq T \). By (8), we have

\[
c(t)x(t)x(\sigma(t))y^\Delta(t) = c(T)x(T)x(\sigma(T))y^\Delta(T) + \int_T^t f(\tau)x(\sigma(\tau))d\tau,
\]

so (19) implies

\[
\text{(21)} \quad \liminf_{t \to \infty} c(t)x(t)x(\sigma(t))y^\Delta(t) = -\infty, \quad \limsup_{t \to \infty} c(t)x(t)x(\sigma(t))y^\Delta(t) = \infty.
\]
By the first relation in (21), there exists \( \tilde{T} \geq T \) with 
\[
\frac{c(\tilde{T})x(\tilde{T})x(\sigma(\tilde{T}))y^{\Delta}(\tilde{T})}{\Delta s} < -2M.
\]
Thus, using (8), we find
\[
c(t)x(t)x(\sigma(t))y^{\Delta}(t) = c(\tilde{T})x(\tilde{T})x(\sigma(\tilde{T}))y^{\Delta}(\tilde{T}) + \int_{\tilde{T}}^{t} f(\tau)x(\sigma(\tau))\Delta \tau
\]
and hence
\[
y(t) < y(\tilde{T}) - 2M \int_{\tilde{T}}^{t} \frac{\Delta s}{c(s)x(s)x(\sigma(s))} + \int_{\tilde{T}}^{t} \frac{1}{c(s)x(s)x(\sigma(s))} \int_{\tilde{T}}^{s} f(\tau)x(\sigma(\tau))\Delta \tau \Delta s
\]
\[
\leq y(\tilde{T}) - M \int_{\tilde{T}}^{t} \frac{\Delta s}{c(s)x(s)x(\sigma(s))} \xrightarrow{(18)} -\infty \quad \text{as} \quad t \to \infty.
\]
By the second relation in (21), there exists \( \bar{T} \geq T \) with 
\[
\frac{c(\bar{T})x(\bar{T})x(\sigma(\bar{T}))y^{\Delta}(\bar{T})}{\Delta s} > 2M.
\]
Thus, using (8), we find
\[
c(t)x(t)x(\sigma(t))y^{\Delta}(t) = c(\bar{T})x(\bar{T})x(\sigma(\bar{T}))y^{\Delta}(\bar{T}) + \int_{\bar{T}}^{t} f(\tau)x(\sigma(\tau))\Delta \tau
\]
and hence
\[
y(t) > y(\bar{T}) + 2M \int_{\bar{T}}^{t} \frac{\Delta s}{c(s)x(s)x(\sigma(s))} + \int_{\bar{T}}^{t} \frac{1}{c(s)x(s)x(\sigma(s))} \int_{\bar{T}}^{s} f(\tau)x(\sigma(\tau))\Delta \tau \Delta s
\]
\[
\geq y(\bar{T}) + M \int_{\bar{T}}^{t} \frac{\Delta s}{c(s)x(s)x(\sigma(s))} \xrightarrow{(18)} \infty \quad \text{as} \quad t \to \infty.
\]
This is a contradiction, as \( y(t) \to \infty \) and \( y(t) \to -\infty \) at the same time for \( t \to \infty \).

**Example 6.** Rankin [1979, Example 2] stated that Theorem 5 for the case \( \mathbb{T} = \mathbb{R} \) can be used to show that
\[
u'' = t \sin t, \quad t \in \mathbb{R}
\]
is oscillatory. Here we let \( x(t) \equiv 1 \). Clearly, conditions (18) and (19) are satisfied.
A simple calculation shows that
\[
\left| \int_{T}^{t} \int_{T}^{s} \tau \sin \tau \cdot d\tau \cdot ds \right| \leq (2T + 4)(t - T),
\]
so (20) is satisfied if $M$ is allowed to depend on $T$. However, if $M = M(T)$, then the proof of Theorem 5 (and Theorem 6) breaks down. Furthermore, the equation (22) is in fact not oscillatory: Clearly,

$$u_1(t) = 4 - 2 \cos(t) + t(2 - \sin t) \quad \text{and} \quad u_2(t) = -t \sin t - 2 \cos t$$

both are solutions of (22), and $u_1$ is nonoscillatory while $u_2$ is oscillatory.

**Example 7.** We note that Grace and El-Morshedy [1997] did not supply an example to illustrate Theorem 5 for the case $\mathbb{T} = \mathbb{Z}$. Consider the difference equation

(23) \[ \Delta^2 u = (-1)^{t+1}(2t + 1), \quad t \in \mathbb{N}. \]

Here we let $x(t) \equiv 1$. Clearly, conditions (18) and (19) are satisfied. A simple calculation shows that

$$\left| \sum_{s=T}^{t-1} \sum_{\tau=T}^{s-1} (-1)^{t+1}(2\tau + 1) \right| \leq 2T(t - T)$$

for $T \in \mathbb{N}$, so (20) is satisfied if $M$ is allowed to depend on $T$. Furthermore, the equation (22) is in fact not oscillatory: Clearly,

$$u_1(t) = t + (-1)^{t+1} \left[ t \right] \quad \text{and} \quad u_2(t) = (-1)^{t+1} \left[ t \right],$$

where $[x]$ denotes the largest integer less than or equal to $x \in \mathbb{R}$, both are solutions of (22), and $u_1$ is nonoscillatory while $u_2$ is oscillatory.

We now present the following results.

**Theorem 7.** Let $T \in \mathbb{T}$. Assume $x$ is any solution of (1) with $c(t)x(t)x(\sigma(t)) > 0$ for all $t \geq T$. If (18) holds and if there exists $M > 0$ such that (20) is satisfied, then (2) is not oscillatory.

**Proof.** Define

$$y(t) := 2M \int_T^t \frac{\Delta x}{c(s)x(s)x(\sigma(s))} + \int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau))\Delta \tau \Delta s.$$  

Using the product and the quotient rule (4), it is easy to check that $u$ defined by $u := yx$ is a solution of (2). However, (20) ensures that $\lim_{t \to \infty} y(t) = \infty$, and therefore $u$ is a nonoscillatory solution of (2). Thus (2) cannot be oscillatory. □

**Theorem 8.** Let $T \in \mathbb{T}$. Assume $x$ is any solution of (1) with $c(t)x(t)x(\sigma(t)) > 0$ for all $t \geq T$. If (18) holds, if there exists $M > 0$ such that (20) is satisfied, and if (11) and (12) hold, then (2) has both oscillatory and nonoscillatory solutions.
Proof. Define

\[ y_1(t) := 2M \int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))} + \int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau)) \Delta\tau \Delta s \]

and

\[ y_2(t) := \int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau)) \Delta\tau \Delta s. \]

As in the proof of Theorem 7, it is easy to check that \( u_1 \) and \( u_2 \) defined by \( u_1 := y_1 x \) and \( u_2 := y_2 x \) both are solutions of (2). While \( u_1 \) is nonoscillatory, \( u_2 \) is oscillatory. Hence (2) indeed has both oscillatory and nonoscillatory solutions. \( \square \)

Our next result can be checked easily as in the proof of Theorem 7.

**Theorem 9.** Suppose that the solution \( x \) of (1) satisfies \( x(t) \neq 0 \) for all \( t \geq T \). Then the solution of (2) satisfying the initial conditions \( u(T) = \alpha \) and \( u^\Delta(T) = \beta \) is given by \( u := y x \), where

\[ y(t) = \gamma + \delta \int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))} + \int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau)) \Delta\tau \Delta s, \]

where \( \gamma = \alpha/x(T) \) and \( \delta = c(T)(\beta x(T) - \alpha x^\Delta(T)) \).

The following generalization of Theorem 2 now becomes apparent.

**Theorem 10.** Suppose \( x \) is a nonoscillatory solution of (1). If for some \( T \in \mathbb{T} \),

\[ \liminf_{t \to \infty} \frac{\int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau)) \Delta\tau \Delta s}{\int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))}} = -\infty \]

and

\[ \limsup_{t \to \infty} \frac{\int_T^t \frac{1}{c(s)x(s)x(\sigma(s))} \int_T^s f(\tau)x(\sigma(\tau)) \Delta\tau \Delta s}{\int_T^t \frac{\Delta s}{c(s)x(s)x(\sigma(s))}} = \infty, \]

then (2) is oscillatory.

\[ \text{References} \]


Received June 7, 2005.

**Martin Bohner**  
Department of Mathematics  
University of Missouri  
Rolla, MO 65401  
United States  
bohner@umr.edu

**Christopher C. Tisdell**  
School of Mathematics and Statistics  
The University of New South Wales  
Sydney, NSW 2052  
Australia  
cct@maths.unsw.edu.au
CURVES ON NORMAL RATIONAL CUBIC SURFACES

JOHN BREVIK

Let $k$ be an algebraically closed field, let $X_0$ be a rational normal cubic surface in $\mathbb{P}^3 = \mathbb{P}^3_k$, and let $C_0 \subset X_0$ be a locally Cohen–Macaulay curve, which is therefore an effective Weil divisor on $X_0$. I show that $C_0$ can be expressed as the limit of a family of curves whose general member lies on a smooth surface, in the following sense: There exists a flat family $X_t$ of cubic surfaces specializing to $X_0$ and a flat family $C_t$ of curves specializing to $C_0$, parametrized by a smooth (noncomplete) curve $T$, such that the general member of $X_t$ is a smooth cubic surface and $C_t \subset X_t$ is an effective (Cartier) divisor for all $t \in T \setminus \{0\}$.

Introduction

Let $k$ be an algebraically closed field. In this paper, I will use the term curve to refer to a locally Cohen–Macaulay scheme of pure dimension 1 over $k$; a space curve is such a curve embedded as a subscheme of $\mathbb{P}^3 = \mathbb{P}^3_k$. As has been frequently observed, even if one’s primary interest is in smooth space curves, one needs to work in the more general category of space curves as defined above in order to use the modern ideas of linkage and minimal curves, since this category is the closure of the category of smooth space curves under liaison (see, for example of [Hartshorne 1980, Section 4] for a discussion of this point).

A common approach to studying space curves is to study the curves on a particular surface or class of surfaces. For example, Gruson and Peskine [1982] constructed a quartic surface with a double line and showed that for any degree-genus pair in a certain range there is a smooth curve with this degree and genus on this quartic, thus answering the question which $(d, g)$ are possible for smooth space curves. Phrased in terms of Hilbert schemes, any $H^d_g$ which contains a smooth curve contains a curve lying on a plane, a smooth quadric surface, a smooth cubic surface, or the quartic with a double line. Mori [1984] then showed that in fact any $(d, g)$ pair that occurs for a smooth curve on this quartic also occurs for a smooth curve on a smooth quartic; this surface, however, depends on the pair $(d, g)$.

MSC2000: primary 14H50; secondary 14J17.

Keywords: classification of space curves, curves on singular surfaces, singular del Pezzo surfaces.
There is also the finer question of which irreducible components of the Hilbert scheme have curves on a given surface or on a member of a class of surfaces. There is little work in the literature in this direction, but one can ask a number of interesting questions along these lines:

(1) For a given curve $C$ on a given singular surface $S$ of degree $s$, is it true that, in the Hilbert scheme, $C$ is in the closure of the space of curves on smooth surfaces of degree $s$?

(2) For a given singular surface $S$, is it true that all curves on $S$ are in the closure of the set of curves on smooth surfaces as above?

(3) If one stratifies in some way the space of all surfaces of a given degree, surfaces in which strata contain the most general curves in some component of some Hilbert scheme?

Here I investigate these questions for $S$ a rational normal surface of degree 3. That the questions are not trivial is shown by a few examples. Consider:

**Theorem** [Hartshorne 1997]. Let $S_0 \subset \mathbb{P}^3$ be a quadric cone. There exists a family of smooth quadric surfaces $S_t$ parametrized by a smooth curve $T$ and specializing to $S_0$ such that every effective Weil divisor $C_0$ on $S_0$ is the limit of a flat family of curves $C_t$, where $C_t \subset S_t$ for all $t \in T$.

Thus, at least some singular surfaces satisfy the very strong property (2) above.

On the other hand, if one allows a quadric surface to degenerate into the union of two planes, there are many curves on this union whose $(d, g)$ pairs do not correspond to the values which arise from curves on a smooth quadric, so certainly not every surface satisfies property (2). Even if we insist that our special surface be integral, it may not satisfy this property: Gruson and Peskine [1982] have produced an example of a family of curves of degree 13 and genus 18, lying on a cubic scroll with a double line, for which they show by dimension-counting that the general member cannot be the specialization of a flat family of curves on smooth cubic surfaces. In [Brevik and Mordasini 2003] we gave very strong necessary conditions for a curve to be such a specialization; a consequence of these conditions is a generalization of Gruson and Peskine’s example: In fact, no smooth curve of degree greater than 10 on this surface is the specialization of a flat family of curves on smooth cubic surfaces.

The present work answers the question (2) above (and therefore question (1) as well) in the affirmative for curves on rational normal cubic surfaces in $\mathbb{P}^3$. In fact, the result is stronger, since it shows that the necessary deformation of surfaces is independent of the curve chosen.

**Theorem.** Let $X_0 \subset \mathbb{P}^3$ be a rational normal cubic surface. Then there exists a (not necessarily complete) curve $T$ with point 0 and a flat family $X \to T$ of cubic...
surfaces in $\mathbb{P}^3$ such that $X_0$ is the fibre over 0, $X_t$ is smooth for $t \neq 0$, and such that any effective Weil divisor $C_0$ on $X_0$ is the limit of a flat family of curves $C$ over $T$ with $C_t \subset X_t$ for each $t \in T$.

With respect to question (3) above, then, the most general member of an irreducible component of a Hilbert scheme is never contained in a rational normal cubic surface and not a smooth cubic. This result is again a generalization of the work of Gruson and Peskine, in the following sense: In the proof of [Gruson and Peskine 1982, Théorème 2.11], root systems are used to show that any $(d, g)$ pair for a smooth curve on a rational normal cubic surface already occurs as a $(d', g)$ pair for a smooth curve on a smooth cubic surface. The method of proof and the current paper’s elementary Proposition 5.19 together imply the result for smooth curves (at least in characteristic 0). The difficulty, then, is to construct an appropriate flat family for general curves in our sense, which may be reducible or nonreduced.

It is natural to ask whether this result generalizes to surfaces of higher degree. Consider, however, the example of a quartic surface $S$ with one $\mathbb{A}_1$ singularity and a line $L$ containing the singular point. Then it is easy to show that the Weil divisor $2L$ on $S$ has arithmetic genus $-2$, and there simply are no curves on a smooth quartic of degree 2 and genus $-2$, since the arithmetic genus of a double line on a smooth quartic is always $-3$. On the other hand, Proposition 5.5 below shows that, for any rational double point surface singularity in $\mathbb{P}^3$, the problem of lifting a divisor to a smooth surface is at least “locally unobstructed,” in some sense, relative to the combinatorics of the exceptional curves on the desingularization. I believe then, that it is natural to conjecture that, given a surface of degree $d$ in $\mathbb{P}^3$ whose only singularities are rational double points, any sufficiently positive (maybe numerically effective) divisor can be expressed as the flat limit of a family whose general member lies on a smooth surface of degree $d$.

Section 1 of this paper briefly recalls some relevant facts about ideals defining 0-dimensional subschemes of a smooth surface and establishes the terminology regarding these objects that will be used in this paper. Section 2 concerns the construction, from the standpoint of blowing up the plane, of a normal cubic surface in $\mathbb{P}^3$. Section 3 “relativizes” the blowing-up process; this technique is used in Section 4 to construct families of surfaces. Section 5 consists largely of a study of numerical invariants of curves on cubic surfaces and culminates in the proof of the main theorem.

1. **Complete ideals and base loci**

This material can be found in [Lipman 1969; 1988]; what it does for us is to establish a dictionary between base conditions on linear systems on a smooth surface and
0-dimensional subschemes of that surface. For a thorough treatment, specialized to our particular situation, see [O'Sullivan 1996].

For any ideal $\mathfrak{I}$ defining a 0-dimensional subscheme of a smooth surface $Y$, and for any point $P$ on $Y$, define $\text{ord}_P(\mathfrak{I})$ to be the largest power of the maximal ideal $m_P$ in the local ring $\mathcal{O}_P$ which contains $\mathfrak{I}_P$. Also, if $X \to Y$ is obtained by blowing up a finite sequence of closed points, define the transform $\mathfrak{I}^X$ of $\mathfrak{I}$ on $X$ to be the ideal obtained from $\mathfrak{I}\mathcal{O}_Y$ by factoring out the invertible part, so that $\mathfrak{I}^X$ again defines a 0-dimensional subscheme. For any point $Q$ existing on some blown-up surface above $Y$, define $\text{ord}_Q(\mathfrak{I})$ to be $\text{ord}_Q(\mathfrak{I}^X)$, where $X$ is the minimal surface above $Y$ for which $Q$ exists as an ordinary point. Then $\mathfrak{I}$ gives rise to a formal sum of points $\sum_Q \text{ord}_Q(\mathfrak{I}) \cdot Q$, to which I will refer as the base locus of $\mathfrak{I}$. (In [Lipman 1988], this is called the point basis of $\mathfrak{I}$.)

Recall that if $I$ is an ideal of a ring $R$, an element $r$ of $R$ is integral over $I$ if $r$ satisfies a polynomial

$$r^n + a_1 r^{n-1} + \cdots + a_n,$$

where each $a_i \in I$. The completion of $I$ is the set of elements of $R$ which are integral over $I$; we say that $I$ is complete if it contains all elements of $R$ integral over $I$. One can show that the completion of $I$ is a complete ideal, and also that completeness is a local property, so it makes sense to speak of the completion of an ideal sheaf.

**Proposition 1.1** [Lipman 1988, Proposition 1.10]. Let $Y$ be a smooth surface. Two ideals $\mathfrak{I}, \mathfrak{J}$ of finite colength on $Y$ have the same base locus if and only if they have the same completion.

**Theorem 1.2** [Lipman 1988, Theorem 3.1]. Let $Y$ be a smooth surface, and let $\mathfrak{I}$ be a complete ideal sheaf of $Y$ defining a 0-dimensional subscheme $Z$. Then the base locus $\sum_Q \text{ord}_Q(\mathfrak{I}) \cdot Q$ of $\mathfrak{I}$ is a finite sum, and the length of $Z$ is equal to

$$\sum_Q \frac{1}{2}(\text{ord}_Q(\mathfrak{I}))(\text{ord}_Q(\mathfrak{I}) + 1).$$

**Remark 1.2.1.** Thus, the base locus defines a mapping from the set of ideals on $Y$ of finite colength to the free additive monoid on the set of (equivalence classes of) points on surfaces birationally dominating $Y$. The fibres are ideals having the same completion. The image is the set of point sums satisfying the proximity inequalities (compare [Zariski 1971, Chapter II, §2]). The proximity inequalities are the set of conditions that no point in a sum has a coefficient which is smaller than the sum of the coefficients of the points infinitely near to it.

Therefore, there is a one-to-one correspondence between the set of complete ideals of finite colength on a surface $Y$ and the set of formal sums of points on surfaces birationally dominating $Y$ which satisfy the proximity inequalities.
Proposition 1.3 [Lipman 1988, Lemma 1.11]. Suppose that \( Z \) is a 0-dimensional scheme on the smooth projective surface \( Y \) with base locus \( B \). Let \( P \) be a closed point on \( Y \), and let \( Y_P \xrightarrow{\pi} Y \) be the blowing-up of \( Y \) at \( P \), with exceptional curve \( E \). Then \( \mathcal{I}_Z \cdot c_{Y_P} = d_{E^\text{ord}_P(Z)} \cdot \mathcal{I}_{Z'} \), where \( Z' \) is a 0-dimensional subscheme of \( Y_P \). Then the base locus \( B' \) of \( Z' \) obeys the formula

\[
B = B' + \text{ord}_P(Z) \cdot P,
\]

where we identify infinitely near points on the two surfaces in the obvious way.

As shorthand, I will sometimes refer to a sum of points on \( \mathbb{P}^2 \) which satisfy the proximity inequalities as a base locus without specifying its ideal. I will refer to a sum of points \( P_1 + \cdots + P_n \) such that, for \( i = 2, \ldots, n \), \( P_i \) is infinitely near to \( P_{i-1} \) and to no other \( P_j \) as a tower.

2. Blowing up \( \mathbb{P}^2 \) at six points

In this section, we will establish some preliminaries for studying rational normal cubic surfaces in \( \mathbb{P}^3 \). The point of view is that such a surface arises from the blowing-up of a set of 6 points in \( \mathbb{P}^2 \), with some of the points possibly infinitely near, satisfying generality conditions (Proposition 2.4). In fact, with the exception of the final result (Theorem 2.5), all of the results in this section are easy to establish using means analogous to those for smooth cubics (see, for example, [Hartshorne 1977, V, Section 4]).

Definition 2.1. Let \( V \) be a smooth quasiprojective surface (over \( k \)) and let \( \tilde{V} \rightarrow V \) be a finite composition of blowings-up of closed points \((Q_1, \ldots, Q_n)\). The strict exceptional divisor \( E_i \) on \( \tilde{V} \) is defined to be the strict transform of the exceptional curve obtained when \( Q_i \) is blown up. The total exceptional divisor \( e_i \) is the total transform of the point \( Q_i \).

Remark 2.1.1. In case the sequence \((Q_1, \ldots, Q_n)\) form a union of towers, then there are restrictions on the order in which the points must be blown up, and it is easily seen that the resulting surface \( \tilde{V} \) and the divisors \( E_i \) and \( e_i \) are independent of that order.

Proposition 2.2. Let \( S = \{P_1, \ldots, P_r\} \) be a union of towers in \( \mathbb{P}^2 \) with the points numbered so that the index of any point is greater than the index of its source. Let \( X \) be the surface obtained by sequentially blowing up the points of \( S \), and let \( E_i \) and \( e_i \) respectively be the strict and total exceptional divisors on \( X \). Let \( D \) be a divisor on \( X \).

1. If \( P_i \) is the source of some other point \( P_j \), then \( E_i = e_i - e_j \).
2. \( \text{Pic } X \cong \mathbb{Z}^7 \), generated by \( \ell, e_1, \ldots, e_6 \).
\begin{enumerate}
\item The intersection pairing on $X$ is given by $\ell^2 = 1$, $\ell \cdot e_i = 0$, $e_i \cdot e_j = -\delta_{ij}$.
\item The canonical class of $X$ is $K = -3\ell + \sum e_i$.
\item The self-intersection of $D$ is $D^2 = a^2 - \sum b_i^2$.
\item If $D$ is effective, its arithmetic genus is $p_a(D) = \frac{1}{2}(D^2 + D.K) + 1 = \left(\frac{a-1}{2}\right) - \sum \left(\frac{b_i}{2}\right)$.
\end{enumerate}

**Proposition 2.3.** Let $S$ be a union of towers consisting of six points $P_1, \ldots, P_6$ of $\mathbb{P}^2$, no four collinear. Let $X$ be the surface obtained by blowing up $S$. Then any integral curve with negative self-intersection must be one of the following:

\begin{enumerate}
\item an exceptional curve $E_i$, which has self-intersection $-2$ if there is a point of $S$ infinitely near to $P_i$ and $-1$ otherwise;
\item the strict transform $F_{ij}$ of a line in $\mathbb{P}^2$ containing only the points $P_i$ and $P_j$ of $S$, which has self-intersection $-1$;
\item the strict transform $F_{ijk}$ of a line in $\mathbb{P}^2$ containing the points $P_i$, $P_j$ and $P_k$ of $S$, which (if it exists) has self-intersection $-2$;
\item the strict transform $G_j$ of a conic in $\mathbb{P}^2$ containing all of $S$ except $P_j$, which has self-intersection $-1$;
\item the strict transform $G$ of a conic in $\mathbb{P}^2$ containing all of $S$, which (if it exists) has self-intersection $-2$.
\end{enumerate}

**Proposition 2.4.** Let $S$ be a union of towers consisting of six points $P_1, \ldots, P_6$ of $\mathbb{P}^2$, no four collinear. Let $X$ be the surface obtained by blowing up $S$. Then the anticanonical divisor $-K$ defines a morphism $\phi$, which is an isomorphism away from $(-2)$-curves and which collapses $(-2)$-curves, from $X$ to a normal rational cubic surface in $\mathbb{P}^3$ having only rational double points as singularities.

Conversely:

**Theorem 2.5** [Nagata 1960, Theorem 8]. Let $X$ be a normal cubic surface in $\mathbb{P}^3$ which is not a cone. Then the minimal desingularization $\tilde{X}$ of $X$ is isomorphic to $\mathbb{P}^2$ blown up at a set $S$ of 6 points which is a union of towers with no 4 points collinear. The morphism $\tilde{X} \to X \subset \mathbb{P}^3$ is given by the linear system induced by the system of cubics through $S$.

**Remark 2.5.1.** A thorough treatment of Theorem 2.5 can be found in [O’Sullivan 1996]. Briefly, the idea of the proof is as follows: First, it follows easily that the singularities of $X$ are double points. Projecting from a singular point gives a birational map to $\mathbb{P}^2$ whose indeterminacy can be resolved by blowing up at the singular point. One then shows that the blown-up surface has only rational singularities, because these are the only singularities that can dominate a smooth
point [Zarisky and Samuel 1960]. Therefore, the original surface also has only rational singularities and thus can be desingularized by a sequence of blowings-up. The minimal desingularization $\widetilde{X}$ dominates $\mathbb{P}^2$, and its canonical divisor has self-intersection 3, since resolving rational double points does not affect the canonical divisor. Thus $\widetilde{X} \to \mathbb{P}^2$ must involve exactly six intermediate blowings-up. That $S$ is a union of towers with no four points collinear follows from the fact that the general hyperplane section of $X$ is a smooth elliptic curve.

3. Relative constructions

The main result of this section is that the process of blowing-up is compatible with flat families in the sense that, under conditions that are met for our applications, blowing up a flat family of schemes along a flat family of closed subschemes gives another flat family, and that taking a fibre of this family gives an identical result to taking a fibre and then blowing up. This will allow us to construct families of cubic surfaces from families of points in the plane and study limits of families of curves on these particular families of surfaces. A few algebraic preliminaries are first necessary.

The following lemma and its “sheafified” sequel are readily verified.

Lemma 3.1. Let $A \to B \to C$ be rings such that $B$ and $C$ are both flat over $A$. Let $M$ be an $A$-module. Then for all $i > 0$,

$$\text{Tor}^i_B(B \otimes M, C) = 0.$$ 

Proposition 3.2. Suppose $Z \xrightarrow{g} T$ and $Y \xrightarrow{f} T$ are flat morphisms and $Z \xrightarrow{h} Y$ is a morphism over $T$. Let $\mathcal{L}$ be a quasicoherent sheaf on $T$. Then, for all $i > 0$,

$$\mathcal{K} \text{Tor}^i_Y(f^*\mathcal{L}, h_*\mathcal{O}_Z) = 0.$$ 

To proceed, we need a result from ring theory. Let $R$ be a ring. Recall that an ideal $I$ of $R$ is regular if it is generated by a regular sequence in $R$; $I$ is perfect if $\text{depth}(I) = \text{pd}_R(R/I)$.

The significance of perfect ideals lies in the following result.

Theorem 3.3 [Balcerzyk and Józefiak 1989, III, Theorem 3.5.11]. Let $R$ be a Cohen–Macaulay ring, $I$ an ideal of $R$ such that $\text{pd}_R(R/I)$ is finite. Then $I$ is perfect if and only if $R/I$ is Cohen–Macaulay.

Theorem 3.4 (Auslander–Buchsbaum; see [Balcerzyk and Józefiak 1989, III, Theorem 3.5.6]). Let $R$ be a local ring with maximal ideal $m$, and let $M$ be a nonzero finitely-generated $R$-module of finite projective dimension. Then

$$\text{pd}_R(M) + \text{depth}(m; M) = \text{depth}(m).$$
Corollary 3.5 [Balcerzyk and Józefiak 1989, III, Theorem 3.5.10]. Let \((R, m)\) be a local ring, \(I\) a regular ideal in \(R\). Then for all \(n \geq 1\), \(I^n\) is a perfect ideal.

The following standard result will be used several times in the sequel.

Proposition 3.6 [Hartshorne 1977, III.9.7]. Let \(X \xrightarrow{f} Y\) be a morphism of schemes, with \(Y\) integral and regular of dimension 1. Then \(f\) is flat if and only if every associated point \(x \in X\) maps to the generic point of \(Y\).

Corollary 3.7. Let \(X\) be a Cohen–Macaulay scheme, \(Z\) a closed subscheme which is a local complete intersection on \(X\), that is, the ideal sheaf \(\mathcal{I} = \mathcal{I}_Z/X\) is locally generated by \(\text{codim}(Z, X)\) elements.

1. For each positive integers \(n\), let \(Z_n\) be the subscheme of \(X\) defined by \(\mathcal{I}^n\). Then \(Z_n\) is a Cohen–Macaulay scheme.
2. Suppose further that \(Z\) is flat over a nonsingular curve \(T\). Then for all \(n\), \(Z_n\) is flat over \(T\).

Proof. (1) For any point \(x \in X\), \((\mathcal{I}_Z)_x\) is a regular ideal in a Cohen–Macaulay local ring, so Corollary 3.5 can be used, which gives the result.

(2) By Proposition 3.6, flatness over \(T\) is equivalent to having all associated points map to the generic point of \(T\). Since \(Z_n\) is Cohen–Macaulay from part (1), it has no embedded points, so the associated points are just the generic points of its components. But these are the components of \(Z\), and so the fact that all of its generic points map to the generic point of \(T\) forces the same to hold true for \(Z_n\). □

Proposition 3.8. Let \(T\) be a smooth curve over \(k\), and let \(Y\) be an integral Cohen–Macaulay scheme which is flat and of finite type over \(T\). Let \(Z\) be a closed subscheme of \(Y\) such that \(Z\) is a local complete intersection on \(Y\) and is also flat over \(T\). Let \(\widetilde{Y}\) be the blowing-up of \(Y\) along \(Z\) with exceptional divisor \(E\). Then \(\widetilde{Y}\) and \(E\) are both flat over \(T\), and for all \(t \in T\), \((\widetilde{Y})_t\) is the blowing-up of \(Y_t\) at \(Z_t\).

Proof. The blow-up map is birational, and \(\widetilde{Y}\) is integral, so by Proposition 3.6, it is flat over \(T\).

Now, let \(Z_n\) be the subscheme of \(Y\) defined by \(\mathcal{J}_Z^n\) and tensor the exact sequence

\[0 \to \mathcal{J}_Z^n \to \mathcal{O}_Y \to \mathcal{O}_{Z_n} \to 0\]

by \(\mathcal{O}_{Y_t}\). Since \(Z_n\) is flat by Corollary 3.7, we can apply Proposition 3.2 to \(\mathcal{L} = k(t)\) so that \(\text{Tor}_i^\mathcal{O}_Y(\mathcal{O}_{Y_t}, \mathcal{O}_{Z_n}) = 0\) for \(i > 0\). Hence we obtain the exact sequence

\[0 \to \mathcal{J}_Z^n \otimes \mathcal{O}_{Y_t} \to \mathcal{O}_{Y_t} \to \mathcal{O}_{Z_n} \otimes \mathcal{O}_{Y_t} \to 0.\]

Let \(\mathcal{J} = \mathcal{J}_Z\mathcal{O}_Y\). A priori, \(\mathcal{J}^n\) is the image of \(\mathcal{J}_Z^n \otimes \mathcal{O}_{Y_t}\) in \(\mathcal{O}_{Y_t}\). But by the injectivity of the map on the left it follows that \(\mathcal{J}_Z^n \otimes \mathcal{O}_{Y_t} \cong \mathcal{J}^n\). This shows that

\[\text{Proj} \bigoplus \mathcal{J}^n \cong \text{Proj}(\mathcal{O}_{Y_t} \otimes \bigoplus \mathcal{J}_Z^n);\]
that is, blowing up is compatible with taking fibres. This is the last statement of
the proposition.

Finally, since $E$ is of pure codimension 1 in $\tilde{Y}$ and is of codimension 1 on every
fibre of $\tilde{Y}$, $E$ can have no components which lie over closed points of $T$. Thus, by
Proposition 3.6, $E$ is flat over $T$. □

4. Families of blown-up surfaces

We will now apply the results of the last section to a configuration of 6 points
in $\mathbb{P}^2$, possibly infinitely near. For this section, we will be considering flat fami-
lies of surfaces over a smooth curve, where a “smooth curve” is understood to be
nonsingular, connected and of finite type over $k$ but not necessarily complete. The
reason that it is important not to insist on completeness for our base schemes is that
eventually (Proposition 4.11) it may become necessary to excise a finite number
of closed points in order to make the results go through.

Proposition 4.1. Let $T$ be a (possibly nonprojective) smooth connected curve over
$k$. Let $Z$ be a flat family of length-$n$, 0-dimensional schemes parametrized by $T$
such that, for all $t \in T$, the fibre $Z_t$ is reduced. Then there exists a surjective base
extension $T' \to T$ of curves such that $Z' = Z \times_T T'$ is the scheme-theoretic union
of $n$ disjoint families of points parametrized by $T'$.

Proof. Since each fibre of $Z$ is a union of reduced points, $Z$ is flat and unramified
over $T$, so (by [Hartshorne 1977, III, Ex. 9.4]) $Z$ is smooth over $T$. Therefore $Z$
is the disjoint union of a number of smooth curves $Z_i$. Let $K$ be the function field
of $T$, and let $K_i$ be the function field of $Z_i$.

Let $K'_1$ be a splitting field for $K_1$ over $K$. Then tensoring the field extension
$K \to K_1$ with $K'_1$ gives $K'_1 \to (K'_1)^{m_1}$; therefore, if $T'_i$ is the normalization of $T$ in
$K'_1$, $Z_1 \times_T T'_1 \cong \bigcup_{i=1}^{m_1} T'_1$; that is, the base-extended family is the union of $m_1$ disjoint
families of points. Now proceed in this manner to the other $Z_i$, finally producing
a curve $T' = T'_n$ over which all of the $Z_i$ have been separated into disjoint families
of points. □

Example 4.1.1. Let $R = k[x, y^2 - (x^3 - x)]$ and $S = (R[y]/y^2 - (x^3 - x))$, so the family
of pairs of points given by the map $R \to S$ corresponds to the unramified part of a
map of degree 2 from an elliptic curve to $\mathbb{P}^1$. (In this case, of course, the extension
of function fields is already Galois and the curve normal.) Then tensoring by $S$
gives $S \to S[z]/z^2 - y^2 \cong S \oplus S$; we have thus base-extended the family to two
disjoint families.

Proposition 4.2. Let $T$ be a smooth curve over $k$, let 0 be a (closed) point of $T$,
and let $Y \to T$ be a flat family of smooth surfaces over $k$. Let $Z \subset Y$ be a family of
length-$n$ schemes such that $Z_t$ is a union of $n$ distinct points for $t \neq 0$. Then there
exists a base extension \( T' \to T \) such that \( Z' = Z \times T' \) is equal to the union of \( n \) smooth irreducible components.

**Proof.** The base extension \( T' \to T \setminus \{0\} \) exists by Proposition 4.1. To make the base extension to all of \( T \), consider the map of complete nonsingular curves induced by the base extension \( T' \to T \setminus \{0\} \) and augment \( T' \) by one of the preimages of 0. \( \square \)

**Definition 4.3.** A 0-dimensional subscheme \( Z \) of a surface \( Y \) is *curvilinear* if \( Z \) is contained locally in a smooth curve on \( Y \).

**Proposition 4.4.** Let \( T \) be a smooth curve over \( k \) with special point 0, and let \( Y \to T \) be a flat family of smooth surfaces over \( k \). Let \( Z \subset Y \) be a family of length-\( n \) schemes such that \( Z_t \) is reduced for \( t \neq 0 \) and such that the scheme \( Z_0 \) is curvilinear with base locus \( S_0 \). Then \( S_0 \) is a union of towers, and the fibre \( Y_0 \) is isomorphic to the sequential blowing-up of \( Y \) at the points of the base locus \( S_0 \).

**Proof.** \( S_0 \) is a union of towers, since any other base locus (satisfying the proximity inequalities) has a coefficient \( \geq 2 \) for some point, which forces a curve containing the base locus to have a multiple point. If necessary, base-extend \( T \) as in Proposition 4.2 so that \( Z \) is the union of \( n \) disjoint copies of \( T \) away from 0. Label the components of \( Z \) as \( P_1, P_2, \ldots, P_n \) and proceed by induction on \( n \):

For \( n = 1 \), we can treat the point 0 just as any other point in Proposition 4.2. Notice that the blowing-up is a flat family of surfaces by Proposition 3.8.

Now let \( Q = (P_1)_0 \), the point at which \( P_1 \) meets \( Y_0 \). Consider the following exact diagram of \( \mathcal{O}_Y \)-modules.

\[
\begin{array}{cccc}
0 & 0 & 0 & 0 \\
\downarrow & \downarrow & \downarrow & \\
0 & \mathcal{F} & \mathcal{I}_{Y_0,Y} & \mathcal{I}_{Z_0,Z} \\
\downarrow & \downarrow & \downarrow & \\
0 & \mathcal{I}_Z & \mathcal{O}_Y & \mathcal{O}_Z \\
\downarrow & \downarrow & \downarrow & \\
0 & \mathcal{I}_{Z_0,Y_0} & \mathcal{O}_{Y_0} & \mathcal{O}_{Z_0} \\
\downarrow & \downarrow & \downarrow & \\
0 & 0 & 0 & 0
\end{array}
\]

Choose \( \mathcal{O}(1) \) very ample on \( Y \) and take \( m \gg 0 \) such that the general element of \( H^0\mathcal{I}_{Z_0,Y_0}(m) \) is smooth and \( H^1\mathcal{F}(m) = 0 \). Let \( s \in H^0\mathcal{I}_Z(m) \) restrict to a smooth element of \( H^0\mathcal{I}_{Z_0,Y_0}(m) \). Then \( s \) is smooth at \( Q \), since \( Y_0 \) is locally defined by a generator of the maximal ideal (coming from the uniformizing parameter at 0) in a regular local ring. Therefore \( s \) defines a locally smooth surface \( W \) in \( Y \) such that \( Z \subset W \). Blow up \( Y \) at \( P_1 \) to obtain \( Y^1 \). On \( Y^1 \), the strict transform \( \tilde{W} \) of \( W \) contains the strict transforms of \( P_2, \ldots, P_n \). Further, the surface \( (Y^1)_0 \) is the blowing-up of \( Y \) at \( Q \) by Proposition 3.8, and the ideal of \( \tilde{W}_0 \) is contained in the transform
$\mathcal{X}_Z^1$ of $\mathcal{X}_Z$, since each ideal differs from its total transform by exactly one copy of the exceptional divisor. $\mathcal{X}_Z^1$ is a complete ideal corresponding to the base locus $S_0 \setminus Q$ on $Y^1$. Now apply induction: $X \rightarrow Y^1$ has the property that $X_0 \rightarrow (Y^1)_0$ is isomorphic to the sequential blowing-up of the points of $S_0 \setminus \{Q\}$, and $(Y^1)_0 \rightarrow Y_0$ is the sequential blowing-up of $Q$. □

Remark 4.4.1. Note that the families $P_i$ can be blown up in any order at all, but the points $Q_i$ of $S_0$ have conditions imposed on the order in which they can be blown up, namely that “lower” points on any tower must be blown up before “higher” ones.

Remark 4.4.2. If $Q_i$ is in some tower with successors $Q_i, \ldots, Q_j$, then $e_i = E_i + E_i + \cdots + E_i$, since at each successive blowing-up, the divisor will pick up exactly one copy of the new exceptional curve.

Proposition 4.5. Let $T$ be a nonsingular curve (as usual, not necessarily complete) over $k$ with point 0, let $Y \rightarrow T$ be a flat family of smooth projective surfaces over $k$. Let $Z \subset Y$ be a flat family of curvilinear length-$n$ subschemes of $Y$, such that $Z_t$ is a union of $n$ distinct points for $t \neq 0$. Suppose further that $Z = \bigcup_{\sigma \in \Sigma} P_\sigma$, where $\Sigma$ is a set of cardinality $n$ and each $P_\sigma \cong T$. Let $S_0$ be the base locus of $Z_0$, which consists of $n$ points with coefficient 1 by Proposition 4.4. Then any choice of numbering $P_1, \ldots, P_n$ of the $P_\sigma$ establishes a unique numbering $Q_1, \ldots, Q_n$ of the points of $S_0$ such that

1. For all $i$ the fibre over 0 of the scheme $Y^1$ obtained by blowing up $P_1, \ldots, P_i$ sequentially is isomorphic to the sequential blowing-up of $Y_0$ at $Q_1, \ldots, Q_i$;
2. If $Q_i = s(Q_j)$, then $i < j$;
3. $Q_i$ is an ordinary point of $Y^i_0$ and $Q_i = P_i \cap Y^i_0$;
4. On $X = Y^n$, the surface obtained by sequentially blowing up the $P_i$, the restriction of the strict transform of $P_i$ to $X_0$ is the total exceptional divisor $e_i$ corresponding to the point $Q_i$.

Proof. For each $i$ in sequence, take $Q_i$ to be the lowest (possibly infinitely near) point not yet blown up on the tower supported at the (closed) point where $P_i$ meets $Y_0$. This is clearly the only ordering that could possibly satisfy the desired conditions. By definition, the second is satisfied, and by Proposition 4.4, the first and third are satisfied.

For the last item, since neither of the divisors in question is going to be influenced by families of points that do not become infinitely near to $Q_i$, we may assume that $Q_1, \ldots, Q_n$ form a tower. Further, each divisor will consist only of divisors from subsequent points, so we may as well assume that $i = 1$. If $n = 1$, there is really nothing to prove, so assume that $n > 1$. As in Proposition 4.4, on
the blow-up $Y^1$ the families $P_2, P_3, \ldots, P_n$ all meet the exceptional divisor $E_1$ at the (ordinary) point $Q_2$ on $Y^1_0$. Therefore when we blow up $P_2$ to get the surface $Y^2$, the divisor $E_1$ gets the point $Q_2$ blown up, so that the strict transform of $E_1$ meets $E_2$ in the new exceptional line above 0. The remaining $P_i$ all pass through a particular point on this line, so on blowing up $P_3$, we see that the strict transform of $E_1$ picks up yet another line above 0. Similarly, each $P_i$ contributes a line, so that on the final surface $X = Y^n$ we will have $(e_1)_0 = E_1 + \cdots + E_n = e_1$ on the surface $X_0$.

**Example 4.5.1.** For an illustration, let $T = \mathbb{A}^1 = \text{Spec} \ k[t]$ and let $Z$ consist of two families of points in $\mathbb{A}^2$ given respectively by $(t, 0)$ and $(0, t)$. To find the ideal for $Z$, note that away from $t = 0$ it is given by the product of the ideals of the families. This product is the ideal

$$(x - t, y)(y - t, x) = (xy, y^2 - ty, x^2 - tx, xy - tx - ty + t^2)$$

$$= (xy, y^2 - ty, x^2 - tx, tx - ty + t^2)$$

$$= (xy, y^2 - ty, x^2 - tx, t(t - x - y)).$$

To find the ideal of the flat family, then, we need to throw into the ideal any elements killed by $t$, in this case $t - x - y$. Therefore the ideal for $Z_0$ is $I_0 = (x^2, xy, y^2, x + y) = (x^2, x + y)$, and the associated base locus is $Q_1 + Q_2$, where $Q_1$ is the origin and $Q_2$ is the tangent direction associated to the line $x + y = 0$.

Now, blowing up the ideal $(x - t, y)$ gives (for the appropriate open affine) the ring map

$$k[x, y, t] \xrightarrow{y^\leftrightarrow w(x-t)} k[x, w, t].$$

The total transform of the family $(x, y - t)$ on this new surface is given locally by the ideal

$$(w(x - t) - t, x) = (t(w + 1), x)$$

so that the strict transform is given by $(x, w + 1)$. Therefore this second family meets the fibre over $t = 0$ at the point on the exceptional curve $x = 0$ where $w + 1 = 0$, corresponding to the tangent direction $x + y = 0$. Therefore, blowing up in the order we did established the correspondence

$$(x - t, y) \leftrightarrow Q_1, (x, y - t) \leftrightarrow Q_2.$$

But by the symmetry of the variables, it is clear that blowing up the points in the other order would have reversed the correspondences.

Having blown up $(x - t, y)$, we now blow up the strict transform $(x, w + 1)$ of the second family. This gives (again on the interesting open affine) the ring map

$$k[x, w, t] \xrightarrow{x \mapsto r(w+1)} k[r, w, t].$$
The composite blowing-up on the special fibre looked like
\[
\begin{array}{ccc}
k[x, y] & \xrightarrow{y \mapsto wx} & k[x, w] \\
& \xrightarrow{x \mapsto r(w + 1)} & k[r, w],
\end{array}
\]
which is the blowing-up of the origin followed by the blowing-up of the tangent direction \(x + y\). Note that the surfaces over \(t = 0\) obtained by blowing up the two families of points in the other order are isomorphic, but that this isomorphism does not extend over any open neighborhood of 0.

We can calculate the divisor class group of the threefold \(X\) obtained from the blowing-up of a family of planes at a family of points by using the following, which is an adaptation of [Hartshorne 1997, Proposition 1.1].

**Proposition 4.6.** Let \(T\) be an irreducible nonsingular curve over \(k\) (again not necessarily complete). Let \(X \to T\) be a projective flat family of surfaces over \(T\) such that \(H^1(\mathcal{O}_X) = 0\) and \(X_t\) is smooth for all \(t \in T\). If \(1 \in T\) is any closed point, then there is an exact sequence
\[
0 \to \text{Pic}\ T \xrightarrow{f^*} \text{Pic}\ X \xrightarrow{\rho} \text{Pic}\ X_1,
\]
where \(X_1\) denotes the fibre over 1 and \(\rho\) is the restriction map.

**Proof.** By Stein factorization [Hartshorne 1977, III, Corollary 11.5], the map \(f\) factors as the composition
\[
X \xrightarrow{f'} Y = \text{Spec} \ f_*\mathcal{O}_X \xrightarrow{g} T,
\]
where \(f'\) has connected fibres and \(g\) is finite. Since \(f\) has connected fibres, this forces the fibres of \(g\) to be connected as well, so that \(g\) must be an isomorphism. Therefore \(f_*\mathcal{O}_X = \mathcal{O}_T\). Then by the Projection Formula [Hartshorne 1977, II, Exercise 5.1], \(f_*f^*\mathcal{L} = \mathcal{L}\) for any invertible sheaf \(\mathcal{L}\) on \(T\). This shows that the map \(f_*\) on Pic is injective.

For exactness in the middle, suppose that \(\mathcal{F}\) is an invertible sheaf on \(X\) and that \(\mathcal{F}_1 = \mathcal{F} \otimes \mathcal{O}_{X_1} \cong \mathcal{O}_{X_1}\). Since then \(H^1(\mathcal{F}_1) = 0\), the theorem on cohomology and base change [Hartshorne 1977, III, Theorem 12.11] shows that \(R^1f_*\mathcal{F} = 0\) in a neighborhood of 1, \(f_*\mathcal{F}\) is locally free in a neighborhood of 1, and \((f_*\mathcal{F}) \otimes k(1) \cong H^0(\mathcal{F}_1) = k\). Hence \(f_*\mathcal{F}\) is invertible in a neighborhood of 1.

Consider the natural map \(f^*f_*\mathcal{F} \to \mathcal{F}\). This map restricts to an isomorphism on the fibre \(X_1\), and it is therefore an isomorphism on a neighborhood of \(X_1\). By properness, there is a neighborhood \(V\) of 1 in \(T\) such that \(f^*f_*\mathcal{F} \cong \mathcal{F}\) on \(f^{-1}(V)\), and hence \(\mathcal{F}\) is in the image of Pic \(V\) over this open set.
Finally, let $P$ be any point of $T \setminus V$, and let $T' = V \cup P, X' = f^{-1}T'$. By [Hartshorne 1977, II, 6.5], there is an exact sequence

$$\mathbb{Z} \to \text{Pic} X' \to \text{Pic} f^{-1}V \to 0$$

where the first map sends 1 to the class of $X_P$ in $\text{Pic} X'$. Since $X_P = f^*(P)$, and the kernel of $\text{Pic} X' \to \text{Pic} X_1$ differs from the kernel of $\text{Pic} f^{-1}(V) \to \text{Pic} X_1$ by a multiple of $X_P$, this shows that any $\mathbb{T}$ on $X'$ which maps to 0 in $\text{Pic} X_1$ comes from $f^*$. Proceeding in this manner on the other points of $T \setminus V$ gives the result. □

**Proposition 4.7.** Let $Z$ be a flat family of length-$n$ schemes in $\mathbb{P}^2_T$ such that $Z_t$ is the union of $n$ distinct points for $t \neq 0$. By Proposition 4.2, we can replace $T$ by a base extension so that $Z$ is the union of $n$ smooth irreducible components. Number these components $P_1, \ldots, P_n$ and let $X$ the resulting family of blown-up surfaces. By Proposition 4.5, the points of the base locus associated to $Z_0$ have an ordering $Q_1, \ldots, Q_n$ imposed on them by the numbering of the $P_i$.

1. $\text{Pic} X / \text{Pic} T$ is a free abelian group with generators $\ell_X, (e_1)_X, \ldots, (e_n)_X$, where $\ell_X = \pi^*c_{\mathbb{P}^2_T}(1)$ and $(e_i)_X$ is the strict transform of $P_i$.

2. Denote by $\ell_t$ and $(e_i)_t$, respectively, the total transforms of $c_{\mathbb{P}^2_T}(1)$ and the strict exceptional curves $e_i$ on the surface $X_t$. Then for each $t \in T$, the natural restriction map $\text{Pic} X \to \text{Pic} X_t$ factors to give an isomorphism $\text{Pic} X / \text{Pic} T \to \text{Pic} X_t$ taking $\ell_X$ to $\ell_t$ and $(e_i)_X$ to $(e_i)_t$.

**Proof.** First, the divisor $(e_i)_X$ restricts to $(e_i)_t$ on each $X_t$ by Proposition 4.5; clearly $\ell_X$ restricts to $\ell_t$ as well. Proposition 4.6 shows that the restriction map gives an injection $\text{Pic} X / \text{Pic} T \to \text{Pic} X_t$ for each fibre, but these maps are actually surjective, since $\ell$ and the $e_i$ generate the Picard groups for the blown-up surfaces. □

**Remark 4.7.1.** As with a single blown-up surface, we will write a divisor class $a \ell - b_1e_1 - \cdots - b_ne_n$ on $X$ as the $n$-tuple $(a; b_1, \ldots, b_n)$.

**Definition 4.8.** I will refer to a length-6 subscheme of $\mathbb{P}^2$ as **general** if it is reduced, does not lie on a conic, and meets no line in length $\geq 3$. Such a subscheme is **almost general** if it is curvilinear and meets no line in length $\geq 4$.

**Remark 4.8.1.** By [Hartshorne 1977, V, Section 4], we see that general subschemes are those whose base loci give smooth cubics when blown up.

**Theorem 4.9** [Fogarty 1968, Theorem 2.4]. Let $d$ be a positive integer. Then the family of 0-dimensional subschemes of $\mathbb{P}^2$ of length $d$ is irreducible.

**Corollary 4.10.** Let $S$ be any 6-point base locus on $\mathbb{P}^2$ which is a union of towers such that no four points of $S$ are contained in a line, and let $Z$ be the associated scheme. Then there exists a smooth curve $T$ with point 0 and a flat family of schemes $Z_T$ over $T$ such that $Z_t$ is general for $t \neq 0$ and $Z = Z_0$. 
Proposition 4.11. Let $T$ be a smooth curve, as usual not necessarily complete. Let

$$Z = P_1 \cup \cdots \cup P_6 \subset \mathbb{P}_T^2, \quad P_i \cong T,$$

be a family over $T$ of length-6 subschemes of $\mathbb{P}^2$ such that $Z_t$ is general for $t \neq 0$ and $Z_0$ is almost general. Let $\tilde{X}$ be the scheme obtained by blowing up $\mathbb{P}_T^2$ sequentially at the curves $P_1, \ldots, P_6$. Then (possibly after shrinking $T$ to a neighborhood of 0) the morphism

$$\tilde{X} \xrightarrow{\phi} \mathbb{P}_T^3$$

given by the complete linear system $(3; 1, 1, 1, 1, 1)$ has image $Y$ which is a flat family of normal cubic surfaces in $\mathbb{P}_T^3$ such that for each $t \neq 0$, $Y_t$ is isomorphic to the image in $\mathbb{P}^3$ of the surface obtained by sequentially blowing up the base locus associated to $Z_t$ (which is just a reduced set of points for $t \neq 0$).

Proof. Let $\mathcal{L}$ be the invertible sheaf on $\tilde{X}$ corresponding to the divisor class $(3; 1, 1, 1, 1, 1)$. First, I claim that $H^1(\mathcal{L}_0) = 0$. To see this, let $H$ be a smooth effective divisor in $|\mathcal{L}_0|$ on $\tilde{X}_0$ (these exist since the image of $\tilde{X}_0$ under $\mathcal{L}_0$ has only isolated singularities) and consider the exact sequence of sheaves on $\tilde{X}_0$

$$0 \to \mathcal{O}_{\tilde{X}_0} \to \mathcal{L}_0 \to \mathcal{O}_H(\mathcal{L}_0) \to 0.$$

Now, $H^1(\mathcal{O}_H) = 0$, and $H^1(\mathcal{O}_H(\mathcal{L}_0)) = 0$ because $\mathcal{L}_0|_H$ is a divisor of degree 3 on an elliptic curve. This forces $H^1(\mathcal{L}_0) = 0$.

Now, again by cohomology and base change, $f_*\mathcal{L}$ is free (of rank 4) in some neighborhood of 0; therefore, shrinking $T$ if necessary, we may assume that there are global sections of $\mathcal{L}$ which restrict to a generating set of $\mathcal{L}_t$ on $\tilde{X}_t$ for each $t \in T$. Then $\mathcal{L}$ is generated by global sections, since at any stalk a nongenerator restricts to a nongenerator on its fibre.

Thus $\mathcal{L}$ gives a morphism $\phi : \tilde{X} \to \mathbb{P}_T^3$. Since the restriction $\mathcal{L}_t$ of $\mathcal{L}$ is the linear system $(3; 1, 1, 1, 1, 1)$ on $\tilde{X}_t$, it gives the map from $\tilde{X}_t$ into $\mathbb{P}^3$ with image the normal cubic surface which comes from the blowing-up of $Z_t$. \qed

Remark 4.11.1. It is via this morphism $\phi$ that we can extract information about divisors on the normal cubic surface $X_0$ from information about the family $\tilde{X}$ and Proposition 4.7. Roughly speaking, one would like to take a divisor $C_0$ on $X_0$, translate it into a divisor $\tilde{C}_0$ on $\tilde{X}_0$, identify the divisor type, and use this information to express $C_0$ as the limit of a flat family of curves on smooth cubic surfaces. The difficulty with this method is that the actual limit of the flat family so obtained may have embedded points, so the challenge is to guarantee somehow that one can find a divisor class on $\tilde{X}_0$ giving $C_0$ such that the limit of the resulting family is without embedded points.
Consider the family $\tilde{X}$ of surfaces obtained from the blowing-up of $\mathbb{P}^2_T$ at the family $Z = P_1 \cup \cdots \cup P_6$, where the $P_i$ are ordinary and general except that $(P_1)$ and $(P_2)$ meet in $\mathbb{P}^2_{k(0)}$, and $(P_3)$ and $(P_4)$ meet at another point of $\mathbb{P}^2_{k(0)}$, so that the base locus $Q_1, \ldots, Q_6$ associated to $Z_0$ (in the sense of Proposition 4.5) has $Q_2$ infinitely near to $Q_1$ and $Q_4$ infinitely near to $Q_3$. Upon sequentially blowing up, we find that $\tilde{X}_0$ has two $(-2)$-curves which do not meet, so that in the family $X \subset \mathbb{P}^3_T$, $X_t$ is smooth for $t \neq 0$ and $X_0$ has two isolated singular points. (In fact, each is a rational double point of type $A_1$, analytically isomorphic to the “cone singularity” $xy - z^2$.) Consider the line $F_{13} \subset \mathbb{P}^2_{k(0)}$ containing $Q_1$ and $Q_3$. Its strict transform in $\tilde{X}_0$ will be mapped under the restriction of $\phi$ to a line $L_0$ in $\mathbb{P}^3$, since it is a limit of the strict transforms of the lines $(1; 1, 0, 1, 0, 0, 0)$ on the smooth cubic surfaces $X_t$. Notice that $L_0$ contains both of the singular points of $X_0$.

Now consider the multiplicity-2 structure $2L_0$ on $L_0 \subset X_0$, defined by throwing out embedded points from the subscheme of $X_0$ defined by $g^2_{L_0, X_0}$. I claim that $2L_0$ is actually contained in a plane. To see this, choose any point of $L_0$ besides the two singularities of $X_0$ and consider the plane section of $X_0$ containing $L_0$ and the normal direction at the chosen point. This plane section has 3 singular points lying on $L_0$, and it is easily shown that any plane curve of degree 3 having 3 singular points along a line must contain the 2-structure on that line, so our chosen plane section contains $2L_0$.

Note that each of the families $F_{13}, F_{14}, F_{23}, F_{24}$ on the smooth cubics specialize to $L_0$, where $F_{ij}$ is the family of lines on the smooth cubics coming from lines through $P_i$ and $P_j$ in the plane. Therefore, there are a number of ways to try to write $2L_0$ as a (flat) limit, and all of them will agree with $2L_0$ generically. However, for example, the limit of $F_{13} + F_{14}$ has an embedded point, since the fibres are disjoint for $t \neq 0$, making the arithmetic genus $-1$, so $2L_0$ is not the flat limit of this family. On the other hand, $F_{13}$ and $F_{24}$ meet in every fibre, so the arithmetic genus of $F_{13} + F_{24}$ is 0, which is the arithmetic genus of $2L_0$. By [Hartshorne 1977, Theorem III.9.7], $L_0$ is the flat limit of the family $F_{13} + F_{24}$.

**Proposition 4.12.** Let $Y$ be a family of smooth surfaces over a nonsingular curve $T$ with point $0 \in T$. Let $L$ be a divisor class on $Y$ whose restriction $L_0$ to $Y_0$ is represented by an effective divisor $D_0$ such that either

1. $H^1(Y_0, L_0) = 0$ or
2. $H^2(Y_0, L_0) = 0$ and $H^1(Y_t, L_t)$ is constant in a neighborhood of 0.

Then, possibly after shrinking $T$ to an open neighborhood of 0, the divisor class $L$ is represented by an effective divisor $D$ on $Y$ without vertical components whose restriction to $Y_0$ is equal to $D_0$. 

The following is needed for the vanishing of certain cohomology groups.

**Lemma 4.13.** Let \( X \) be a smooth surface, and let \( C \subseteq X \) be a reduced Cartier divisor such that each irreducible component of \( C \) has negative intersection with the canonical divisor \( K_{\tilde{\mathbb{P}}} \) of \( \tilde{\mathbb{P}} \). Then \( H^1(C, C) = 0 \).

**Proof.** Let \( \omega_C \) be the dualizing sheaf of \( C \), and let \( \omega_X \) be the canonical sheaf of \( X \). Then \( \omega_C \cong \omega_X|_C \). By Serre duality, \( H^1(C, C) \perp H^0(\omega_C(-C)) = H^0(\omega_X|_C) \). Therefore it suffices to show that \( H^0(\omega_X|_C) = 0 \).

First, assume that \( C \) is integral. Then \( \omega_X|_C \) is a Cartier divisor of negative degree on the integral curve \( C \), so \( H^0(\omega_X|_C) = 0 \).

Now suppose that \( C = \bigcup_{i=1}^r C_i \), where the \( C_i \) are integral. Let \( Y \) be the disjoint union of the \( C_i \). Then the natural projection \( f : Y \rightarrow C \) exhibits \( \omega_X|_C \) as a subsheaf of \( f^*(\bigoplus \omega_X|_{C_i}) \). Since \( H^0(\bigoplus \omega_X|_{C_i}) \cong H^0(f^*(\bigoplus \omega_X|_{C_i})) \) (see [Hartshorne 1977, III, Ex. 8.2], for example), and since \( H^0(\omega_X|_{C_i}) = 0 \) for all \( i \) by the previous paragraph, \( H^0(\omega_X|_C) = 0 \) and the lemma is proved. \( \square \)

### 5. The Main Theorem

**Configuration of \((-2\)-curves on the blown-up surface.** In this section, I will identify the possible configurations of \((-2\)-curves on a surface obtained by blowing up a 6-point base locus in \( \mathbb{P}^2 \) which is a union of towers with no four base points collinear.

**Definition 5.1.** I will refer to a connected set of curves as a **cluster**. If a surface singularity arises from the collapsing of a cluster of \((-2\)-curves to a point, this singularity is sometimes labelled according to the Dynkin diagram of the configuration of the \((-2\)-curves; see Definition 5.1. One can show that the configuration of such \((-2\)-curves is an invariant of the analytic isomorphism class of the singularity (see [Lipman 1969] or [O’ Sullivan 1996] for details). It is well-known that the intersection matrix of the set of \((-2\)-curves collapsing to a point is negative-definite (see [Artin 1962]).

It is known [Artin 1966] that any rational double point surface singularity contained in \( \mathbb{P}^3 \) is of type \( A_r, D_r, E_6, E_7, \) or \( E_8 \); for rational normal cubic surfaces, this list is shorter, namely \( A_1, \ldots, A_5, D_4, D_5, \) or \( E_6 \) (see [Bruce and Wall 1979]).

Recall from [Artin 1966] that the **fundamental cycle** \( \xi_0 \) associated to a rational surface singularity is defined (locally) as the smallest exceptionally supported positive cycle on the minimal desingularization with the property that its intersection with each exceptional curve is nonpositive. Fundamental cycles for the rational double points above are calculated as follows. Note that I express each in terms of its coefficients in the free monoid on the exceptional curves \( E_i \); thus, for example,
Figure 1. Top to bottom: Dynkin diagrams for $A_r$, $D_r$, and $E_r$ configurations. The circles in the diagrams stand for curves, and a connecting line segment between two circles indicates that the corresponding curves intersect transversally at a single point.

$(2, 3, 1)$ stands for the divisor $2E_1 + 3E_2 + E_3$. The curves are numbered as in Figure 1.

- For $A_r$, $\xi_0 = (1, 1, \ldots, 1)$.
- For $D_r$, $\xi_0 = (1, 2, 2, 2, \ldots, 2, 1, 1)$.
- For $E_6$, $\xi_0 = (1, 2, 3, 2, 2, 1)$.
- For $E_7$, $\xi_0 = (1, 2, 3, 4, 2, 3, 2)$.
- For $E_8$, $\xi_0 = (2, 3, 4, 5, 6, 3, 4, 2)$.

**Definition 5.2.** Suppose that $\widetilde{S}$ is a surface arising as the minimal resolution of a rational normal cubic surface in $\mathbb{P}^3$, and let $D$ be an effective divisor on $\widetilde{S}$. $D$ has **small intersection with fundamental cycles** if, for each connected cluster $\{C_1, \ldots, C_r\}$ of $(-2)$-curves on $\widetilde{S}$ with fundamental cycle $\xi_0$, $D$ has nonnegative intersection multiplicity with each of the $E_j$ and $D.\xi_0 \leq 1$. (Equivalently, $D$ has nonzero intersection multiplicity with at most one of the curves in the cluster, and furthermore, if there is such a curve $C_k$, then $D.E_k = 1$ and $C_k$ has coefficient 1 in $\xi_0$.)
Proposition 5.3. Let \((R, \mathfrak{m})\) be a local ring defining a rational double point on a surface, let \(\text{Spec} \, S \xrightarrow{\pi} \text{Spec} \, R\) be the minimal desingularization, with fundamental cycle \(\xi_0\), and let \(D \subseteq \text{Spec} \, S\) be a nonexceptional smooth curve. Then the multiplicity of the scheme-theoretic image \(\pi(D)\) at \(\mathfrak{m}\) is equal to \(D.\xi_0\).

Proof. Artin [1966, proof of Theorem 4] shows that \(\mathcal{J}(\xi_0) = \mathfrak{m} \cdot S\). Now, restricting the map \(R \to S\) to \(D\) gives a map from the local ring \((A, q)\) of the scheme-theoretic image of \(D\) on \(\text{Spec} \, R\) to a semilocal ring \(B\) whose maximal ideals \(p_i\) correspond to the points of \(D\) on the exceptional divisor. Now, each localization \(B_{p_i}\) is a discrete valuation ring, and the multiplicity \(e_i\) of \(p_i\) in the primary decomposition of \(\mathfrak{m}B\) is precisely the coefficient of the component of the exceptional curve containing the closed point \(p_i\) in \(\xi_0\), from Artin’s above observation. Now, apply Corollary 1 to [Zariski and Samuel 1960, vol. II, VII, Theorem 24] to conclude that

\[ m_{\mathfrak{m}}(\pi(D)) = \sum m_{p_i}(D) \cdot e_i. \]

From the preceding considerations, this last sum is equal to \(D.\xi_0\). \(\square\)

Corollary 5.4. Let \(\widetilde{\mathbb{P}}_S \xrightarrow{\pi} S\) be the minimal resolution of a rational normal cubic surface in \(\mathbb{P}^3\), and let \(D\) be an effective divisor on \(\widetilde{\mathbb{P}}_S\) having small intersection with fundamental cycles. Then \(D \to \pi(D)\) is an isomorphism.

Proof. Locally, the map on \(D\) is a birational morphism of smooth curves. \(\square\)

Proposition 5.5. Let \(S\) be a 6-point base locus in \(\mathbb{P}^2\) which is a union of towers such that no four points of \(S\) are collinear. Let \(\widetilde{\mathbb{P}}_S \xrightarrow{\pi} S\) be the surface obtained by blowing up \(S\) sequentially. Let \(D\) be an effective divisor on \(\widetilde{\mathbb{P}}_S\) having nonnegative intersection with fundamental cycles. Then there exists an effective divisor \(E\), supported on \((-2)\)-curves, such that \(D + E\) has small intersection with fundamental cycles.

Proof. Clearly, it suffices to consider each connected cluster of \((-2)\)-curves separately. Let \(\{C_1, \ldots, C_r\}\) be such a cluster, and let \(a_i = D.C_i\). When convenient, I will adopt the notation using ordered \(r\)-tuples for exceptionally supported divisors introduced above. I will also use square brackets to indicate intersection multiplicities with the \(-2\)-curves; thus, I would write that \(D\) has intersection multiplicities \([a_1, \ldots, a_r]\).

First, suppose that the \(C_i\) form an \(A_r\), \(r \geq 1\). Proceed by induction on \(s = \Sigma a_i\). If \(s \leq 1\), then \(D\) has small intersection with the fundamental cycle, and we are done.

If \(s \geq 2\), first suppose that there are two indices \(i < j\) such that \(a_i\) and \(a_j\) are positive. Then add the divisor \(C_i + C_{i+1} + \cdots + C_j\). This addition reduces \(a_i\) and \(a_j\) by 1 each and leaves the other intersection numbers unchanged; by induction, we are finished.
Finally, if there is only one index $i$ such that $a_i$ is positive, then $a_i \geq 2$. Add a
single copy of $E_i$ to $D$. Then there are two possibilities: If $1 \neq i \neq r$, then there
are at least two entries, namely the $(i-1)^{st}$ and the $(i+1)^{st}$, in the new divisor
which are nonzero, and we may proceed as in the last paragraph; if $i = 1$ or $i = r$,
then the sum of the entries has dropped, since the $i$-th has dropped by 2 and there
is only one which has increased by 1, and we are finished by induction.

For a $D_4$, note that $\xi_0.C_i = -\delta_i2$. By adding copies of $C_1$, $C_3$, and $C_4$ to $D$,
we may assume that each of $a_1$, $a_3$, $a_4$ is less than or equal to 1; and by adding a
multiple of $\xi_0$, we may assume that $a_2 = 0$. By symmetry, it suffices to assume
that $a_1 \geq a_3 \geq a_4$.

If $a_3 = 0$, then $D$ has small intersection with the fundamental cycle.

If $a_1 = a_3 = 1$, $a_4 = 0$, then add the divisor $(1, 1, 1, 0)$ to get a divisor with
intersections $[0, 0, 0, 1]$, which has small intersection with the fundamental cycle.

If $a_1 = a_3 = a_4 = 1$, then adding the divisor $(2, 3, 2, 2)$ gives intersections
$[0, 0, 0, 0]$.

Thus the proposition is true for a $D_4$.

For $r > 4$, observe that the curves $C_2, \ldots, C_n$ form a $D_{r-1}$, so by induction
we may assume that there is an divisor $L$, supported on $C_2 \cup C_3 \cup \cdots \cup C_n$, such
that $D' = D + L$ has intersection number 1 with at most one of $C_2$, $C_{r-1}$, and $C_r$
and intersection number 0 with all other $C_i$ except $C_1$. Write the new intersection
sequence as $[b_1, \ldots, b_r]$, and note that $b_1 \geq a_1 \geq 0$. If $b_1 \geq 2$, we can reduce $b_1$
by 2 and leave the other $b_i$ unchanged by adding the divisor $(2, 2, \ldots, 2, 1, 1)$, so
we may assume that $b_1$ is either 0 or 1.

If $b_1 = b_2 = 0$, then we are done, as $D'$ has small intersection with the funda-
mental cycle. If $b_1 = 1$ and all other $b_i$ are 0, we are also done. If $b_1 = b_2 = 1$,
then $D' + E_0$ has small intersection with $\xi_0$. If $b_1 = b_{r-1} = 1$, then
$D' + C_1 + C_2 + \cdots + C_{r-1}$ has small intersection with $\xi_0$. Treat the case $b_1 = b_r = 1$
similarly.

This exhausts all possibilities, so the $D_r$ case is complete.

In the case of an $E_6$, notice that adding the fundamental cycle $\xi_0$ reduces the
intersection number with $C_4$ by 1 and leaves the others unchanged; therefore, we
can always assume that the intersection number of our divisor of interest with $C_4$
is 0. From the $A_r$ case above, add an effective exceptionally supported divisor to
$D$ in order to obtain a divisor $D'$ whose intersection numbers with $C_1$, $C_2$, $C_3$, $C_5$,
and $C_6$ are all 0 except for possibly one of them; if there are none, clearly we are
done, and if $D'.C_1 = 1$ or $D'.C_6 = 1$, we are done as well, since these curves
have multiplicity 1 in $\xi_0$. If $D'.C_2 = 1$, add $(0, 1, 1, 0, 1, 1)$; similarly for $C_5$. If
$D.C_3 = 1$, add $(1, 3, 4, 2, 2, 0)$. 
For an $\mathbf{E}_7$, note that adding the fundamental cycle reduces the intersection number with $C_7$ by 1 and leaves the others unchanged; as above, then, we need not worry about this intersection number. Also, note that $C_1, \ldots, C_6$ form a $\mathbf{D}_6$; therefore, we can find an effective divisor supported on these curves to add to $D$ in order to obtain $D'$, whose sequence of intersection multiplicities is either: all 0’s, in which case we are done; $[1, 0, 0, 0, 0, 0, 0]$, which already has small intersection with the fundamental cycle; $[0, 0, 0, 0, 0, 1, 0]$, in which case we add $(2, 4, 6, 8, 4, 6, 3)$ to obtain all 0’s; or $[0, 0, \ldots, 1, 0, 0]$, in which case we add $(0, 1, 2, 3, 2, 3, 2, 1)$ to get back to $[1, 0, 0, 0, 0, 0, 0]$. 

Finally, in the case of an $\mathbf{E}_8$, adding $\delta_0$ reduces the intersection number with $C_1$ and leaves the others unchanged. $C_2, \ldots, C_8$ form an $\mathbf{E}_7$, so from the previous case we may assume that we can add an appropriate divisor to $D$ and obtain $D'$ whose sequence of intersection multiplicities is either all 0’s (done) or $[0, 1, 0, 0, 0, 0, 0, 0]$. In this latter case, adding $(3, 6, 8, 10, 12, 6, 8, 4)$ reduces the intersection multiplicities to all 0’s. This finishes the proof.

**Definition 5.6.** Label the standard basis elements of $\mathcal{Z}^r$ as $\ell, e_1, \ldots, e_{r−1}$, and denote by $f_{ij}$ the elements $\ell+e_i+e_j$ for $i \neq j$. Define the Cremona transformation $\tau_{ijk}$ for $i, j, k$ all different to be the change of basis $\ell', e'_1, \ldots, e'_n$, where $\ell' = l, e'_i = f_{jk}, e'_j = f_{ik}, e'_k = f_{ij}$, and $e'_n = e_n$ for $n$ different from $i, j, k$.

**Remark 5.6.1.** The reason I call this a Cremona transformation is that it is a generalization of the change of basis induced by the plane Cremona on the surface obtained by blowing up three noncollinear points, which makes sense whether the points are ordinary or infinitely near.

**Lemma 5.7.** Let $D$ be an element of $\mathcal{Z}^7$. Then, by a finite number of applications of the Cremona transformation $\tau_{123}$ and reorderings, one can produce a basis $\ell, e_1, \ldots, e_6$ for $\mathcal{Z}^7$ such that, with respect to this basis, the representation $D = a\ell - \sum b_i e_i$ satisfies

1. $b_1 \geq b_2 \geq b_3 \geq b_4 \geq b_5 \geq b_6$
2. $a - b_1 - b_2 - b_3 \geq 0$.

**Proof.** Label the standard basis elements for $\mathcal{Z}^7$ as $\ell, e_1, e_2, \ldots, e_6$. In order to avoid cumbersome notation, I will abuse notation and use the same names for elements of each basis produced on the way to the desired one. Define a bilinear form on $\mathcal{Z}^7$ as on a smooth cubic surface, $i.e. \ell^2 = 1, \ell.e_i = 0, e_i.e_j = -\delta_{ij}$, and label the elements $f_{ij} = l - e_i - e_j, g_{ij} = 2l - \sum_{l \neq m} e_i$. Then condition (2) is interpreted as $D.f_{12} \geq D.e_3$. In order for the $b_i$ to be in descending order, this forces $D.f_{ij} = a - b_i - b_j \geq D.e_3 = b_3$. Note also that $D.e_1 \geq D.e_2 \geq D.e_3$. Finally, $a - b_i - b_j - b_k \geq 0$ for any $i, j, k$, so $D.g_m = 2a - \sum_{l \neq m} b_i \geq b_3$ for all $g_m$. Therefore, with respect to our desired basis, the elements $e_6, e_5, e_4, e_3$ must have,
in order, the four smallest intersections with $D$ among the set $\mathcal{I} = \{ e_1, f_{ij}, g_k \}$. I will show that an algorithm exists which accomplishes this.

By repeated applications of the Cremona transformation $\tau_{123}$ and reorderings, whatever element of $\mathcal{I}$ having minimal intersection with $D$ can be brought to $e_6$. Then the remaining $e_i$ can be renumbered so that the $b_i$ are in descending order. Then, as in the previous paragraph, the element of $\mathcal{I}$ which has the next minimal intersection with $D$ is $e_5$, $f_{12}$, or $g_6$.

If it is $e_5$, proceed to the next step in the algorithm.

If it is $f_{12}$, applying $\tau_{123}$ leaves $e_4$, $e_5$, and $e_6$ alone but takes $f_{12}$ to $e_3$. By renumbering, then, $e_5$ can be made to have next smallest intersection with $D$.

Finally, if it is $g_6$, apply $\tau_{123}$ so that $g_6$ becomes $f_{45}$; by renumbering, this will then become $f_{12}$, and we are in the previous case.

So far we have the desired $e_6$ and $e_5$. Again renumber the remaining $e_i$ so that the $b_i$ are descending. At this point, a priori again the candidates for the next smallest intersection with $D$ are $e_4$, $f_{12}$, and $g_6$, but I claim that we need not consider $g_6$. To see this, note that by choice of $e_5$, $D.e_5 \leq D.f_{12}$, or in other words $b_5 \leq a - b_1 - b_2$.

Substituting this inequality gives

$$D.g_6 = 2a - b_1 - b_2 - b_3 - b_4 - b_5$$

$$\geq 2a - b_1 - b_2 - b_3 - b_4 - (a - b_1 - b_2)$$

$$= a - b_3 - b_4$$

$$\geq a - b_1 - b_2$$

$$= D.f_{12}.$$ 

Therefore, the next element of our basis is either $e_4$, in which case we can move on to the next step, or $f_{12}$, which is treated just like the second case in the step above and does not affect our $e_5$ or $e_6$.

The final step in the algorithm goes just like the previous one, since the Cremona transformation $\tau_{123}$ does not affect $e_5$, $e_6$, or $e_4$, so finally I have produced a basis with the desired properties.

\[\Box\]

**Proposition 5.8.** Let $\overline{\mathbb{P}}$ be a surface resulting from the blowing-up of a 6-point base locus $(P_1, \ldots, P_6)$ in $\mathbb{P}^2$ which is a union of towers such that no four base points are on a line. Let $D$ be a divisor class on $\overline{\mathbb{P}}$ such that the intersection multiplicity of $D$ with any $(-2)$-curve is nonnegative. Then there exists a map $\overline{\mathbb{P}} \to \mathbb{P}^2$ which is again the blowing-up of a (possibly different) 6-point base locus $(P'_1, \ldots, P'_6)$ in $\mathbb{P}^2$ such that no four base points are on a line and such that with respect to the induced basis $\ell, e_1, \ldots, e_6$, $D = (a; b_1, \ldots, b_6)$ with the properties:
If $P_j'$ is infinitely near to $P_i'$, then $i < j$;

- $b_1 \geq b_2 \geq \cdots \geq b_6$;
- $a \geq b_1 + b_2 + b_3$.

**Proof.** If $P_j$ is infinitely near to $P_i$, the condition $D.E_j \geq 0$ forces $D.e_i \geq D.e_j$. Therefore, any renumbering of the points such that the second condition is satisfied will automatically satisfy the first condition as well. With respect to this renumbering, $P_1$ is ordinary, $P_2$ is either ordinary or infinitely near to $P_1$, and $P_3$ is either ordinary or infinitely near to either $P_1$ or $P_2$. Thus $S' = P_1 + P_2 + P_3$ is a union of towers and either $P_1, P_2, P_3$ are collinear or $P_1, P_2, P_3$ determine a Cremona transformation of $\mathbb{P}^2$.

In the first case, $F_{123}$ is a $(-2)$-curve and $D.F_{123} = a - b_1 - b_2 - b_3 \geq 0$, so the last condition holds. Otherwise, the transformation $\tau_{123}$ corresponds to an actual Cremona transformation of $\mathbb{P}^2$, so we can proceed as in Lemma 5.7 to obtain the desired basis. 

**Definition 5.9.** Following the notation in [O’Sullivan 1996], I will refer to the basis determined in Proposition 5.8 for the divisor $D$ as a **preferred basis** for $D$. The 7-tuple of integers representing the divisor class of $D$ with respect to a preferred basis for $D$ is called a **preferred form** for $D$.

**Proposition 5.10.** Let $D = (a; b_1, \ldots, b_6)$ be a divisor class on a smooth cubic surface $X$ with respect to the basis $\ell, e_1, \ldots, e_6$ such that $b_1 \geq b_2 \geq \cdots \geq b_6$ and $a \geq b_1 + b_2 + b_3$. Then $D$ is effective if and only if $a \geq \max(0, b_1)$.

**Proof.** That $a \geq \max(0, b_1)$ is necessary is fairly clear: $a$ cannot be negative for an effective divisor, and $a < b_1$ would mean that there was a plane curve of degree $a$ with a point of multiplicity more than $a$, which is impossible.

To prove sufficiency, let $r = \max\{i : b_i \geq 0\}$ and write $D = D_+ + D_-$, where $D_+ = (a; b_1, \ldots, b_r, 0, \ldots, 0)$ and $D_- = (0; 0, 0, \ldots, b_{r+1}, \ldots, b_6)$. $D_-$ is clearly effective, as it is the sum of exceptional curves. At this point, either $D_+$ satisfies the hypotheses of the proposition or $b_1 < 0$ and $a - b_1 - b_2 = c < 0$. In the latter case, $D_+$ intersects $F_{12}$ with multiplicity $-c$, so it contains at least $c$ copies of $F_{12}$. Now, make the further decomposition $D_+ = D_+ + cF_{12}$ and now $D'_+ \geq 0$ does satisfy the hypotheses of the proposition.

So it suffices to show that the proposition holds for $D$ with all $b_i \geq 0$. First dispensing with an easy case, if $a = b_1$, then the other $b_i$ are all 0 and $D$ is a multiple of the conic $(1; 1, 0, 0, 0, 0, 0)$.

Therefore assume that $a > b_1$. By [Hartshorne 1977, V, Ex. 4.8], if a divisor on a smooth cubic intersects each of the 27 lines nonnegatively and has nonnegative self-intersection, then that divisor is linearly equivalent to an effective divisor. The intersection with the lines is immediate: Since all of the entries for $D$ are positive,
Let $D.E_i \geq 0$ for each $i$. Also, $D.F_{ij} \geq D.F_{12}$ by the condition $b_1 \geq b_2 \geq \cdots \geq b_6$, and $D.F_{12} = a - b_1 - b_2 \geq 0$ by $a \geq b_1 + b_2 + b_3$, $b_3 \geq 0$. Finally, $D.G_j \geq D.G_6$, again by $b_1 \geq b_2 \geq \cdots \geq b_6$, and since $a \geq b_4 + b_5$, so $D.G_6 = 2a - b_1 - b_2 - b_3 - b_4 - b_5 \geq 0$. $D^2 > 0$ is an immediate consequence of the following lemma, which is readily verified.

**Lemma 5.11.** Let $(a; b_1, b_2, \ldots, b_6)$ be a 7-tuple of nonnegative integers such that $a \geq b_1 + b_2 + b_3$, $a > b_1$, and $b_1 \geq b_2 \geq \cdots \geq b_6$. Then $a^2 > \sum b_i^2$.

**The arithmetic genus of an effective divisor on the blown-up surface.** Let $X$ be a smooth projective surface over $k$. Let $D$ be an effective divisor on $X$. Recall that the Zariski decomposition [1962, §7] of $D$ is the unique decomposition $D = D_L + D_N$ with $D_L$ effective such that

1. $D_L = 0$ or the intersection matrix for each connected component of the support of $D_L$ is negative-definite;
2. $D_N$ is numerically effective;
3. $D_N \cdot E = 0$ for each prime divisor $E$ in the support of $D_L$.

The following is needed for Proposition 5.14, which characterizes the Zariski decomposition of an effective divisor on one of our blown-up surfaces.

**Proposition 5.12.** Let $X$ be a smooth projective surface over $k$. Let $\overline{X} \xrightarrow{\pi} X$ be the composition of blowings-up of a sequence of closed points $(P_1, \ldots, P_n)$. Let $D$ be an exceptionally supported divisor such that $D.E_i \leq 0$ for each component $E_i$ of the exceptional locus. Then $D$ is effective.

**Proof.** This is immediate from the following lemma about bilinear forms, which is not difficult to verify. □

**Lemma 5.13.** Let $(\cdot, \cdot)$ be a negative-definite bilinear form on a finitely generated free abelian group $G$ and let $x_1, \ldots, x_n$ be a basis of $G$ such that $(x_i, x_j) \geq 0$ for $i \neq j$. If $y = \sum s_i x_i$ such that $(y, x_i) \leq 0$ for each $i$, then all of the $s_i$ are nonnegative.

**Proposition 5.14.** Let $\overline{P}$ be the surface obtained by blowing up a curvilinear base locus in $\mathbb{P}^2$ of length 6, and let $D$ be an effective divisor on $\overline{P}$ having nonnegative intersection with all $(-2)$-curves, with Zariski decomposition $D = D_L + D_N$. Then $D_N$ is effective. Further, if $D$ is in preferred form, this decomposition for $D$ has one of the following forms:

1. If $a - b_1 - b_2 \geq 0$, then $D_N$ has type $(a, b_1, \ldots, b_r - 1, 0, \ldots, 0)$ and $D_L$ has type $(0; 0, \ldots, 0, b_r, \ldots, b_6)$, where $r$ is such that $b_{r-1} \geq 0$ and $b_r < 0$.
2. If $a - b_1 - b_2 = c < 0$, then $D_N$ has type $(a + c; b_1 + c, b_2 + c, 0, 0, 0)$ and $D_L$ has type $(-c; -c, -c, b_3, \ldots, b_6)$ with $b_3, \ldots, b_6$ all less than 0.
Proof. First, we need to show that our purported $D_L$ is effective. This is immediate, because in each case $D_L$ is a positive linear combination of $e_i$ and $f_{12} = (1; 1, 1, 0, 0, 0, 0)$, which are always effective regardless of the collinearity or infinitely near behavior of the six points in $\mathbb{P}^2$. Further, in both cases each of these components is either a single $(-1)$-curve or a cluster of $(-2)$-curves together with a $(-1)$-curve all collapsing to a point in $\mathbb{P}^2$, and one can verify that its intersection matrix must be negative-definite.

To show that our $D_N$ is numerically effective, it suffices to show that $D_N.E \geq 0$ for any curve $E$ with negative self-intersection. For the exceptional curves $E_1, \ldots, E_6$, this follows from preferred form 5.8: If $E_i$ is a $(-1)$-curve, then $D_N.E_i$ is just the $i$-th coefficient after the semicolon in the divisor type for $D_N$, which is always nonnegative. If $E_i$ is a $(-2)$-curve, say with $P_j$ infinitely near to $P_i$, then $i < j$; by the decreasing nature of the $b_i$ (which property is preserved when passing to $D_L$) the intersection number $D_L.E_i = D_L.e_i - D_L.e_j$ remains nonnegative. Since $D$ was given in preferred form, it is easy to see that $D_N.F_{ij}$ is at least as large as $D_N.F_{12}$.

To show that $D_N$ has 0 intersection with each component of $D_L$, note that, in case 1, $D_L$ is supported on $E_r, \ldots, E_6$ and the result follows trivially from the intersection theory on $\widetilde{\mathbb{P}}$. For case 2, $D_L$ is supported on $F_{12}$ or $F_{123}$, depending on whether $P_1, P_2, P_3$ are collinear, together with $E_3, \ldots, E_6$. By construction $D_L.F_{12} = 0$ or $D_L.F_{123} = 0$, and as in case 1 $D_L.E_i = 0, i = 3, 4, 5, 6$.

Therefore, the Zariski decomposition is as stated. Effectiveness of $D_N$ now follows immediately from Proposition 5.12, since item (3) in the Zariski decomposition shows that $D_N$ is effective on the support of $D_L$, and away from this support it is effective because the original $D$ is. \hfill \Box

Lemma 5.15. Let $\widetilde{\mathbb{P}}$ be the surface obtained from $\mathbb{P}^2$ by the sequential blowing-up of a curvilinear 6-point base locus with no four base points collinear. Let $|D|$ be a numerically effective divisor class on $\widetilde{\mathbb{P}}$. Then the general member of $|D|$ is a reduced divisor each of whose integral components has negative intersection with the canonical divisor $K_{\widetilde{\mathbb{P}}}$ of $\widetilde{\mathbb{P}}$.

Proof. Since $|D|$ is numerically effective, it has no fixed components by Proposition 5.14. Since there are only finitely many curves with negative self-intersection on $\widetilde{\mathbb{P}}$ by Proposition 2.3, the general member of $|D|$ is supported on integral curves of nonnegative self-intersection. Such curves always move in linear equivalence classes, so the general member of $|D|$ is reduced. As $-K_{\widetilde{\mathbb{P}}}$ is very ample away from $(-2)$-curves, each component of the general member of $|D|$ has positive intersection with $-K_{\widetilde{\mathbb{P}}}$. \hfill \Box

Proposition 5.16. Let $\widetilde{\mathbb{P}}$ be the surface obtained by the sequential blowing-up of a 6-point base locus in $\mathbb{P}^2$ which is a union of towers with no four base points
collinear. Let \( D \subset \mathbb{P} \) be an effective divisor such that its intersection with each \((-2)\)-curve is nonnegative. Write the divisor class of \( D \) as \((a; b_1, \ldots, b_6)\) with \( b_1 \geq b_2 \geq \cdots \geq b_6 \) and \( a \geq b_1 + b_2 + b_3 \) as in Proposition 5.8. Then

\[
h^0\mathcal{O}_D(D) = \left(\frac{a+2}{2}\right) - \sum \left(\frac{b_i+1}{2}\right) + \left(\frac{-c}{2}\right),
\]

where \( c = a - b_1 - b_2 \).

**Proof.** Write \( D = D_N + D_L \) as in Proposition 5.14, where \( D_N \) is numerically effective and effective and \( D_L \) is common to all divisors linearly equivalent to \( D \). Then \( h^0\mathcal{O}_D(D) = h^0\mathcal{O}_D(D_N) \). I claim that the formula also gives the same result for \( D \) and \( D_N \). To see this, first write the divisor type of \( D_N \) as \((a_N; b_{1N}, \ldots, b_{6N})\).

If \( c \geq 0 \), then by Proposition 5.14, \( a_N = a \) and \( b_{iN} \leq b_i \) so that \( c_N = a_N - b_{1N} - b_{2N} \) \( \geq 0 \) as well. Therefore,

\[
\left(\frac{-c}{2}\right) = 0, \quad \left(\frac{-c_N}{2}\right) = 0.
\]

Also, the \( b_i \) which contribute nonzero terms to the formula are exactly those which are greater than or equal to 1, in which case \( b_{iN} = b_i \) so the contributions are the same. If \( b_i < 1 \), so is \( b_{iN} \), so the contributions for these \( i \) are both 0.

If \( c < 0 \), then \( c_N = a - b_1 - b_2 - c = 0 \); furthermore, \( b_i = b_{iN} = 0 \) for \( i = 3, 4, 5, 6 \). Therefore, for \( D \) the formula gives

\[
\left(\frac{a+2}{2}\right) - \left(\frac{b_1+1}{2}\right) + \left(\frac{b_2}{2}\right) + \left(\frac{-c}{2}\right);
\]

note that in order for \( c \) to be negative, \( b_1 \) and \( b_2 \) must both be positive. On the other hand, the formula for \( D_N \) gives

\[
\left(\frac{a+c+2}{2}\right) + \left(\frac{b_1+c+1}{2}\right) + \left(\frac{b_2+c+1}{2}\right) = \left(\frac{2a-b_1-b_2+2}{2}\right) + \left(\frac{a-b_2}{2}\right) + \left(\frac{a-b_1}{2}\right).
\]

One can verify that each of the two sums evaluates to

\[
a^2 + 2a - ab_1 - ab_2 + b_1b_2 - b_1 - b_2 + 1.
\]

In either case, then, the formula gives the same result for \( D \) and for \( D_N \), as claimed.

It therefore suffices to prove the formula for \( D_N \); for ease of notation, replace \( D_N \) by \( D \). Then \( D \) is numerically effective, \( c \geq 0 \), and each \( b_i \) is nonnegative by Proposition 5.14. Therefore, Lemma 5.15 shows that \( D \) is linearly equivalent to a reduced divisor each of whose integral components has negative intersection with \( K_{\mathbb{P}} \); since we are proving a formula about dimensions of linear systems, we may as well assume that \( D \) itself has this form. Consider the exact sequence on \( D \)

\[
0 \to O_D \to \mathcal{O}_D(D) \to \mathcal{F} \to 0,
\]

where \( \mathcal{F} \) is a sheaf of sections of \( D \).
Let \(\mathcal{T}\) be a torsion sheaf of degree \(D^2\). By Lemma 4.13, \(H^1(\mathcal{O}_D(D)) = 0\), so \(H^0(\mathcal{O}_D(D)) = \chi(\mathcal{O}_D(D)) = \chi(\mathcal{T}) + \chi(\mathcal{O}_D) = D^2 + 1 - p_a(D) = \frac{1}{2}(D^2 - D.K)\), the last equality coming from the adjunction formula. Now consider the exact sequence

\[
0 \to \mathcal{O}_D \to \mathcal{O}_D(D) \to \mathcal{T} \to 0
\]

and note that, since \(\mathcal{P}\) is rational, \(H^1(\mathcal{O}_{\mathcal{P}}) = 0\). Therefore \(H^1(\mathcal{O}_{\mathcal{P}}(D)) = 0\), and thus, using Proposition 2.2, we find that

\[
H^0(\mathcal{O}_{\mathcal{P}}(D)) = \chi(\mathcal{O}_{\mathcal{P}}(D)) = \chi(\mathcal{O}_D(D)) + \chi(\mathcal{O}_{\mathcal{P}}) = \frac{1}{2}(D^2 - D.K) + 1 = \frac{1}{2}((a^2 - \sum b_i^2) + (3a - \sum b_i)) + 1 = \left(\frac{a+2}{2}\right) - \sum\left(\frac{b_i+1}{2}\right) + \left(\frac{-c}{2}\right).
\]

\[\square\]

**Corollary 5.17.** Let \(\mathcal{P}\) be a surface obtained by blowing up a 6-point base locus in \(\mathbb{P}^2\) which is a union of towers with no four base points collinear. If an effective divisor \(C\) on \(\mathcal{P}\) has nonnegative intersection with all \(-2\)-curves, then \(C\) has the numerical type of an effective divisor on a smooth cubic surface.

**Corollary 5.18.** Let \(\mathcal{P}\) be the surface obtained from \(\mathbb{P}^2\) by sequentially blowing up a curvilinear 6-point base locus with no four base points collinear. Let \(D \subset \mathcal{P}\) be an effective divisor such that its intersection with each \((-2)\)-curve is nonnegative. Write the divisor class of \(D\) as \((a; b_1, \ldots, b_6)\) with \(b_1 \geq b_2 \geq \cdots \geq b_6\) and \(a \geq b_1 + b_2 + b_3\) as in Proposition 5.8, and suppose further that \(a \geq b_1 + b_2\). Then \(H^1(\mathcal{O}_{\mathcal{P}}(D)) = 0\).

**Proof.** The canonical sheaf of \(\mathcal{P}\) is antieffective, and \(D\) is effective, so by Serre duality \(H^2(\mathcal{O}_{\mathcal{P}}(D)) = 0\). Therefore, it suffices to show that \(\chi(\mathcal{O}_{\mathcal{P}}(D)) = H^0(\mathcal{O}_{\mathcal{P}}(D))\).

Consider the exact sequence

\[
0 \to \mathcal{O}_D \to \mathcal{O}_D(D) \to \mathcal{T} \to 0
\]

where \(\mathcal{T}\) is a torsion sheaf of degree \(D^2\). By definition, \(\chi(\mathcal{O}_D) = 1 - p_a(D)\); therefore,

\[
\chi(\mathcal{O}_D(D)) = 1 - p_a(D) + D^2.
\]

Now use the exact sequence

\[
0 \to \mathcal{O}_D \to \mathcal{O}_D(D) \to \mathcal{O}_D(D) \to 0
\]

together with the fact that \(H^1(\mathcal{O}_{\mathcal{P}}) = H^2(\mathcal{O}_{\mathcal{P}}) = 0\) to see that

\[
\chi(\mathcal{O}_{\mathcal{P}}(D)) = 2 - p_a(D) + D^2.
\]
Using Proposition 2.2, the right hand side can be expanded to
\[ 2 - \binom{a-1}{2} + \sum \binom{b_i}{2} + a^2 - \sum b_i^2 \]
which simplifies to
\[ \binom{a+2}{2} - \sum \binom{b_i+1}{2} \]
Since by hypothesis \( a \geq b_1 + b_2 \), this is equal to the formula for \( H^0(\mathcal{O}_{\tilde{D}}(D)) \) found in Proposition 5.16.

**Proposition 5.19.** Let \( T \) be a smooth connected (not necessarily complete) curve, and let \( Z \) be a flat family of length-6 schemes in \( \mathbb{P}^2_T \) such that \( Z_t \) is general for \( t \neq 0 \) and \( Z_0 \) is curvilinear meeting no line in multiplicity 4. Let \( X \) be the resulting family of blown-up surfaces. Let \( D = (a; b_1, \ldots, b_6) \in \text{Pic} X/\text{Pic} T \) as in Proposition 4.6 such that \( D_0 \) is effective and \( D_0.E \geq 0 \) for all \((-2)\)-curves \( E \subset X_0 \). Then the dimensions of the cohomology groups \( H^0(X_t, \mathcal{O}_{X_t}(D_t)) \) and \( H^1(X_t, \mathcal{O}_{X_t}(D_t)) \) are constant over the family, and \( H^2(X_t, \mathcal{O}_{X_t}(D_t)) = 0 \) for all \( t \in T \).

**Proof.** First, find a preferred basis for \( D_0 \) so that the integers \( a, b_1, \ldots, b_6 \) satisfy the conclusion of in Proposition 5.8. Now, \( D_0 \) satisfies the hypotheses of Proposition 5.16; therefore,
\[ h^0(\mathcal{O}_{X_0}(D_0)) = \binom{a+2}{2} - \sum \binom{b_i+1}{2} + \binom{-c}{2}, \]
where \( c = a - b_1 - b_2 \). For this number to be positive, surely \( a \geq b_1 \), so by Proposition 5.10, \( D_t \) is effective for all \( t \in T \). Therefore, the formula equally well applies to all the \( D_t \), so \( H^0(X_t, \mathcal{O}_{X_t}(D_t)) \) is constant on the family. Also, since the canonical divisor \( K_{X_t} \) is antieffective, by Serre duality shows that \( H^2(X_t, \mathcal{O}_{X_t}(K_{X_t} - D_t)) = 0. \) Now use the Riemann–Roch formula [Hartshorne 1977, V, Theorem 1.6] on the surface \( X_t \) to show that
\[ h^0(X_t, \mathcal{O}_{X_t}(D_t)) - h^1(X_t, \mathcal{O}_{X_t}(D_t)) = \chi(\mathcal{O}_{X_t}(D_t)) = D_t.(K_{X_t} - D_t). \]
Since this last intersection number depends only on the integers \( a, b_1, \ldots, b_6 \), and the \( h^0 \) term was already shown to be constant over the family, the \( h^1 \) must be constant over the family as well.

**Proposition 5.20.** Let \( T \) be a smooth connected (not necessarily complete) curve, and let \( Z \) be a flat family of length-6 schemes in \( \mathbb{P}^2_T \) such that \( Z_t \) is general for \( t \neq 0 \) and \( Z_0 \) is curvilinear meeting no line in multiplicity 4. Let \( X \) be the resulting family of blown-up surfaces, and let \( D_0 \subset X_0 \) be an effective divisor having non-negative intersection with all \((-2)\)-curves. Then, possibly after shrinking \( T \), there exists an effective divisor \( D \) on \( X \) without vertical components whose restriction to \( X_0 \) is equal to \( D_0 \).
Proof. By Proposition 5.19, the family satisfies the hypotheses of Proposition 4.12, which gives the desired result immediately. □

The main theorem. At this point we have shown that any divisor on $X_0$ meeting certain conditions can be realized as the limit of a family of divisors on $X_t$. We will now show that for a given curve on a normal cubic surface, one of these families gives rise to a flat family specializing to the given curve.

The strategy involved will be to show that any curve on a normal cubic surface can be “linked up” to a smooth curve by adding hyperplane sections. Then we will compare genus formulas for our blown-up surfaces with known linkage formulas for surfaces in $\mathbb{P}^3$ and deduce that the special fibre of a suitably-chosen family from the blown-up surfaces will give a special fibre in $\mathbb{P}^3$ that has the right arithmetic genus to make a flat family. The following two propositions are thus needed to proceed.

Proposition 5.21. Let $\widetilde{\mathbb{P}}$ be a surface resulting from the blowing-up of a 6-point base locus in $\mathbb{P}^2$ which is a union of towers such that no four base points are on a line. Let $D$ be a divisor class on $\widetilde{\mathbb{P}}$ such that $D$ has small intersection with $(-2)$-curves. Let $M = -K = (3; 1, 1, 1, 1, 1, 1)$ be the anticanonical divisor class of $\widetilde{\mathbb{P}}$. Then for $n$ sufficiently large, the divisor $D + nM$ is basepoint-free and gives a birational morphism to projective space whose image is a surface with isolated singularities.

Proof. First, by Proposition 5.8 we can find a preferred basis for $D$. Note that this basis is also preferred for $D + nM$, since the conditions for a basis to be preferred are unaffected by the addition or subtraction of $M$. By Proposition 2.4, $M$ is very ample away from $(-2)$-curves, so we may choose $n$ large enough that $D + nM$ is also very ample away from $(-2)$-curves. By Lemma 5.15, we can also choose $n$ large enough that $D + nM$ has no base components. Therefore, for such $n$ the general member of $|D + nM|$ is integral, since it is smooth away from $(-2)$-curves by Bertini’s Theorem and has no $(-2)$-curves as components. If necessesary add on another $M$ so that $(D + nM)M > 1$. I claim that for this value of $n$, $D' = D + nM$ is basepoint-free. It suffices to show that the sheaf $\mathcal{O}_{\widetilde{\mathbb{P}}}(D + nM)$ is generated by global sections at each point of the $(-2)$-curves.

As $D$ has small intersection with fundamental cycles, so does $D + nM$; therefore, for each $(-2)$-curve $E$ on $\widetilde{\mathbb{P}}$, the sheaf $\mathcal{O}_E(D')$ is generated by global sections. Hence it will suffice to show that in the exact sequence

$$0 \rightarrow \mathcal{O}_{\widetilde{\mathbb{P}}}(D'-E) \rightarrow \mathcal{O}_{\widetilde{\mathbb{P}}}(D') \rightarrow \mathcal{O}_E(D') \rightarrow 0$$

the map on the right is surjective on global sections; for this, in turn, it suffices to show that $H^1(\mathcal{O}_{\widetilde{\mathbb{P}}}(D'-E)) = 0$. 


Consider the exact sequence
\[ 0 \to \mathcal{O}_{D'}(D' - E) \to \mathcal{O}_{D'}(D') \to \mathcal{O}_{D' \cap E}(D) \to 0. \]

Since \( D' \) has small intersection with fundamental cycles, \( D' \cap E \) is either empty or a reduced point, so the map on the right is surjective on global sections. On the other hand, by Lemma 4.13, \( H^1(\mathcal{O}_{D'}(D')) = 0 \), so this forces \( H^1(\mathcal{O}_{\tilde{D}'}(D' - E)) = 0 \), as desired.

Therefore, the divisor \( D + n \mathcal{M} \) is basepoint-free and thus determines a morphism from \( \tilde{D}_\mathcal{M} \) to projective space. It is an isomorphism away from \( (-2) \)-curves; therefore, the general hyperplane section is smooth away from the images of the \( (-2) \)-curves. Moreover, the general hyperplane section meets at most one \( (2) \)-curve in each cluster, and it is smooth there as well: In fact by Corollary 5.4 its image in the associated normal cubic surface remains smooth, so the general hyperplane section is smooth. \( \square \)

The particulars of the following formula are not important for our applications, only that this formula exists and depends only on the numerical data of the curves and surfaces involved.

**Proposition 5.22** [Martin-Deschamps and Perrin 1990, Chapter III, Proposition 3.2]. Let \( C \) be a curve of degree \( d \) contained in the surface \( Q \) of degree \( s \) in \( \mathbb{P}^3 \). Let \( H \) be the hyperplane class on \( Q \), and let \( D \) be a curve linearly equivalent to \( C + nH \) for some integer \( n \). Then
\[ p_a(D) = p_a(C) + nd + \frac{1}{2}ns(s + n - 4). \]

**Proposition 5.23.** Let \( X_0 \subset \mathbb{P}^3 \) be a normal cubic surface and \( \tilde{X}_0 \xrightarrow{\pi} X_0 \) its desingularization arising from the blowing-up of a curvilinear length-6 base locus with no four points on a line. Let \( C_0 \subset X_0 \) be a (Cohen–Macaulay) curve. Then there exists a divisor \( \tilde{C}_0 \) on \( \tilde{X}_0 \) such that
(a) \( \pi|_{\tilde{C}_0} : \tilde{C}_0 \to C_0 \) is an isomorphism away from \( (-2) \)-curves;
(b) \( \tilde{C}_0 \) has small intersection with fundamental cycles;
(c) \( p_a(\tilde{C}_0) = p_a(C_0) \).

**Proof:** Let \( \tilde{C}_0 \) be the strict transform of \( C_0 \) on \( \tilde{X}_0 \). Then \( \tilde{C}_0 \) has no \( (-2) \)-curves in its support, so its intersection with each \( (-2) \)-curve on \( \tilde{X}_0 \) is nonnegative. Therefore by Proposition 5.5, we can adjust \( \tilde{C}_0 \) by an effective sum of \( (-2) \)-curves to obtain \( \tilde{C}_0 \) having small intersection with fundamental cycles. This \( \tilde{C}_0 \) satisfies items (a) and (b).

For item (c), let \( \mathcal{M} = -K \) be the divisor class \( (3; 1, 1, 1, 1, 1) \) on \( \tilde{X}_0 \). By Proposition 5.21, for large enough \( n \) the linear system \( |\mathcal{L}(\tilde{C}_0) + n\mathcal{M}| \) is basepoint-free and gives a birational map from \( \tilde{X}_0 \) to a surface in projective space having
isolated singularities. Fix such an $n$. Then the divisor class $\mathcal{L}(\tilde{C}_0) + n\mathcal{M}$ contains a smooth curve $\tilde{D}_0$. As $\mathcal{M}$ has 0 intersection with all $(-2)$-curves (remember that $\mathcal{M}$ gave the map $\pi$, which collapses all $(-2)$-curves), $\tilde{D}_0$ still has small intersection with fundamental cycles. Therefore $\pi$ takes $\tilde{D}_0$ isomorphically onto its image $D_0$ in $X_0$, so certainly $p_a(\tilde{D}_0) = p_a(D_0)$. Also note that $D_0$ is linearly equivalent to $C_0 + nH$ on $X_0$, where $H$ is the hyperplane class.

I claim that this forces $p_a(\tilde{C}_0) = p_a(C_0)$. To see this, note that by Corollary 5.17 there is a curve $C$ on a smooth cubic surface $S$ with the same numerical type as $\tilde{C}_0$, since $\tilde{C}_0$ has nonnegative intersection with $(-2)$-curves. The degree of $C$ in $\mathbb{P}^3$ is equal to the degree of $C_0$, since in each case this is computed by taking the intersection with $(3; 1, 1, 1, 1, 1, 1)$. Further, $p_a(C) = p_a(\tilde{C}_0)$, since the arithmetic genus is computed purely by the numerical type of a divisor on a $\tilde{X}_0$ or on a smooth cubic. Let $D$ be a curve in the class of $C + nH$ on $S$. Again by intersection theory, $p_a(D) = p_a(\tilde{D}_0)$, which is equal to $p_a(D_0)$ from the preceding paragraph, and also $D$ and $D_0$ have the same degree. Therefore, the formula in Proposition 5.22 shows that $p_a(C_0) = p_a(C)$, and thus that $p_a(C_0) = p_a(\tilde{C}_0)$. □

**Theorem 5.24.** Let $X_0 \subset \mathbb{P}^3$ be the rational normal cubic surface associated to the blow-up $\tilde{X}_0$ of $\mathbb{P}^2$ at a set $S_0$ of 6 points which is a union of towers with no four points collinear, corresponding to the subscheme $Z_0$ with complete ideal $\mathfrak{I}_{Z_0}$.

Let $Z \subset \mathbb{P}^2_T$ be a flat family over a (not necessarily complete) smooth curve $T$ with $0 \in T$ such that $Z_t$ is general for $t \neq 0$ and $(Z)_0 = Z_0$. As in Proposition 4.11, this construction gives rise to a family $\tilde{X}$ of smooth surfaces which maps into $\mathbb{P}^3_T$ as a family $X$ of normal cubic surfaces with general member smooth and with special member $X_0$. Then (possibly after shrinking $T$) any effective Weil divisor on $X_0$ is the limit of a flat family of curves on $X_t$.

**Proof.** Let $D_0 \subset X_0$ be any effective Weil divisor. Let $\tilde{D}_0$ be its strict transform on $\tilde{X}_0$ (that is, the sum of the appropriate multiples of the strict transforms of the irreducible components of $D_0$). By the last proposition, there exists a divisor $\tilde{D}_0$ on $\tilde{X}_0$ which has small intersection with fundamental cycles, which has the same arithmetic genus as $C_0$, and such that $\pi|_{\tilde{D}_0} : \tilde{D}_0 \to D_0$ is generically an isomorphism. Then by Proposition 5.20, $\tilde{D}_0$ is the limit of a flat family of divisors on $X$. The image of this family under $\pi$ is a family of curves parametrized by $T$. By [Hartshorne 1977, Theorem III.9.7], this image is a flat family it suffices to check Hilbert polynomials. The degree of $D_0$ is equal to the degrees of each of the $D_t$ by the intersection properties on the smooth surfaces $X_t$. The arithmetic genus is constant since $p_a(D_0) = p_a(\tilde{D}_0)$. Therefore the Hilbert polynomials are constant over the family, and $D_0$ has been expressed as the limit of a flat family, as desired. □
Acknowledgments

Special thanks to Mike O’Sullivan: much of this work arose out of our very fruitful collaboration. Thanks also to Robin Hartshorne, for his guidance, mentorship, and countless suggestions; and to Rich Liebling, Scott Nollett, and Enrico Schlesinger, for many stimulating discussions.

This work was carried out while I was at the Department of Computer Science, University of California, Santa Barbara.

References


JOHN BREVIK
DEPARTMENT OF MATHEMATICS AND STATISTICS
1250 BELLFLOWER BOULEVARD
CALIFORNIA STATE UNIVERSITY
LONG BEACH, CA 90840-1001
UNITED STATES
jbrevik@csulb.edu
IDEAL STRUCTURE OF $C^*$-ALGEBRAS ASSOCIATED WITH $C^*$-CORRESPONDENCES

TAKEshi KATSURA

We study the ideal structure of $C^*$-algebras arising from $C^*$-correspondences. We prove that gauge-invariant ideals of our $C^*$-algebras are parameterized by certain pairs of ideals of original $C^*$-algebras. We show that our $C^*$-algebras have a nice property that should be possessed by a generalization of crossed products. Applications to crossed products by Hilbert $C^*$-bimodules and relative Cuntz–Pimsner algebras are also discussed.

Introduction

For a $C^*$-algebra $A$, a $C^*$-correspondence over $A$ is a (right) Hilbert $A$-module with a left action of $A$. Since endomorphisms (or families of endomorphisms) of $A$ define $C^*$-correspondences over $A$, we can regard $C^*$-correspondences as (multivalued) generalizations of automorphisms or endomorphisms. This point of view has the same philosophy as the idea that certain topological correspondences are generalizations of continuous maps [Katsura 2004a, Section 1].

A crossed product by an automorphism is a $C^*$-algebra which has an original $C^*$-algebra as a $C^*$-subalgebra, and reflects many aspects of the automorphism. For example, the set of ideals of the crossed product that are invariant under the dual action of the one-dimensional torus $\mathbb{T}$ corresponds bijectively to the set of ideals of the original $C^*$-algebra that are invariant under the automorphism. As $C^*$-correspondences are generalizations of endomorphisms, a natural problem is to define “crossed products” by $C^*$-correspondences. There is plenty of evidence that the construction given in [Katsura 2003a] for the $C^*$-algebra $\mathcal{O}_X$ from a $C^*$-correspondence $X$ is the right one. One piece of evidence given there is that this construction generalizes many constructions that were or were not considered as generalizations of crossed products. We are going to explain another piece of

MSC2000: 46L05, 46L55.

Keywords: $C^*$-algebra, $C^*$-correspondence, Hilbert module, Cuntz–Pimsner algebras, ideal structure, crossed products.

The author was supported in part by a Research Fellowship for Young Scientists of the Japan Society for the Promotion of Science.
evidence. For a $C^*$-correspondence $X$, we can naturally define a notion of representations of $X$ (Definition 2.7). Thus one $C^*$-algebra which is naturally associated with a $C^*$-correspondence $X$ is a $C^*$-algebra $\mathcal{T}_X$ having a universal property with respect to representations of $X$ (Definition 3.1). This $C^*$-algebra $\mathcal{T}_X$ is none other than the (augmented) Cuntz–Toeplitz algebra defined in [Pimsner 1997]. When a $C^*$-correspondence $X$ is defined by an automorphism, the $C^*$-algebra $\mathcal{T}_X$ is isomorphic to the Toeplitz extension of the crossed product by the automorphism defined in [Pimsner and Voiculescu 1980]. This $C^*$-algebra is too large to reflect the information in $X$. In order to get crossed products, we have to go to a quotient of $\mathcal{T}_X$. There are two ways to proceed. One is to define the covariance of representations of a $C^*$-correspondence $X$, and define a crossed product by $X$ so that it has the universal property with respect to covariant representations of $X$. This kind of method has been used in many papers, and we define our $C^*$-algebra $\mathcal{O}_X$ along this line (Definitions 3.4 and 3.5). The other way is to list up the properties of $\mathcal{T}_X$ that the crossed product should have, and define a crossed product by $X$ to be the smallest quotient of $\mathcal{T}_X$ among the quotients satisfying these properties. For this method, the following two properties seem to be reasonable:

(i) The original $C^*$-algebra is embedded into the crossed product,

(ii) There exists a “dual action” of $\mathbb{T}$ on the crossed product.

In this paper, we show that these two methods give the same $C^*$-algebra $\mathcal{O}_X$ (Proposition 7.14). This indicates that the $C^*$-algebra $\mathcal{O}_X$ is the right one for a “crossed product” by a $C^*$-correspondence $X$. We note that Cuntz–Pimsner algebras do not satisfy the property (i) above when the left action of the $C^*$-correspondence is not injective, and that the $C^*$-algebra $\mathcal{O}_X$ is isomorphic to the Cuntz–Pimsner algebra when the left action of the $C^*$-correspondence is injective.

The “dual action” of $\mathbb{T}$ on the $C^*$-algebra $\mathcal{O}_X$ is called the gauge action. The main purpose of this paper is to describe the all ideals of the $C^*$-algebra $\mathcal{O}_X$ associated with a $C^*$-correspondence $X$ that are invariant under the gauge action. We define invariance of ideals of $A$ with respect to a $C^*$-correspondence $X$ over $A$ (Definition 4.8). Unlike the case of crossed products by automorphisms, we need extra ideals of $A$ other than invariant ideals to describe all gauge-invariant ideals of $\mathcal{O}_X$. Similar facts were observed in many papers ([Bates et al. 2002; Drinen and Tomforde 2005; Katsura 2003b; Katsura 2006a] to name a few) for $C^*$-algebras arising from graphs or topological graphs. We introduce a notion of $O$-pairs, which are pairs consisting of invariant ideals and extra ideals of $A$, and show that gauge-invariant ideals are parameterized by $O$-pairs (Theorem 8.6).

This paper is organized as follows. In Sections 1 and 2, we fix notation and gather results on Hilbert $C^*$-modules and $C^*$-correspondences. In Section 3, we give the definition of our $C^*$-algebras $\mathcal{O}_X$ constructed from $C^*$-correspondences $X$. 

In Sections 4 and 5, we introduce and study invariance of ideals, $T$-pairs and $O$-pairs. These are related to representations of $C^*$-correspondences. In Section 6, we construct a $C^*$-correspondence $X_\omega$ from a $T$-pair $\omega$, and in Section 7 we prove that this $C^*$-correspondence $X_\omega$ has a certain universal property. As a corollary, we give an alternative definition of our $C^*$-algebras $\mathcal{O}_X$ described above (Proposition 7.14).

In Section 8, we prove the main theorem (Theorem 8.6) which says that the set of all gauge-invariant ideals of $\mathcal{O}_X$ corresponds bijectively to the set of all $O$-pairs of $X$. We also see that a quotient of $\mathcal{O}_X$ by a gauge-invariant ideal falls into the class of our $C^*$-algebras. In Section 9, we see that every gauge-invariant ideals have hereditary and full $C^*$-subalgebras which are isomorphic to $C^*$-algebras associated with $C^*$-correspondences. As a consequence of the study of crossed products by Hilbert $C^*$-bimodules in Section 10, all gauge-invariant ideals themselves are shown to be isomorphic to $C^*$-algebras associated with $C^*$-correspondences. In Section 11, we apply our investigation to the relative Cuntz–Pimsner algebras defined in [Muhly and Solel 1998].

We denote by $\mathbb{N} = \{0, 1, 2, \ldots \}$ the set of natural numbers, and by $\mathbb{C}$ the set of complex numbers. We denote by $\mathbb{T}$ the group consisting of complex numbers whose absolute values are 1. We use a convention that $\gamma(A, B) = \{\gamma(a, b) \in D \mid a \in A, b \in B\}$ for a map $\gamma : A \times B \to D$ such as inner products, multiplications or representations. We denote by $\text{span}\{\cdots\}$ the closure of linear span of $\{\cdots\}$. The Hewitt–Cohen factorization theorem can be stated as follows:

**Lemma.** Let $A$ be a $C^*$-algebra, $X$ be a Banach space, and $\pi : A \to B(X)$ a bounded homomorphism from $A$ to the Banach algebra $B(X)$ of the bounded operators on $X$. Then we have $\pi(A)X = \overline{\text{span}}(\pi(A)X)$.

We use this result just to abbreviate the notation and arguments. Readers not familiar with the theorem may use $\overline{\text{span}}(\pi(A)X)$ instead of $\pi(A)X$; the two spaces are actually the same (for a proof, see [Raeburn and Williams 1998, Proposition 2.33] for example).

**1. Hilbert $C^*$-modules**

**Definition 1.1.** Let $A$ be a $C^*$-algebra. A (right) Hilbert $A$-module $X$ is a linear space with a right action of the $C^*$-algebra $A$ and an $A$-valued inner product $\langle \cdot, \cdot \rangle_X$ satisfying certain conditions such that $X$ is complete with respect to the norm defined by $\|\xi\|_X = \|\langle \xi, \xi \rangle_X \|^{1/2}$ for $\xi \in X$.

For a precise definition of Hilbert $C^*$-modules, consult [Lance 1995]. We do not assume that a Hilbert $A$-module $X$ is full. Thus $\overline{\text{span}}(X, X)_X$ can be a proper ideal of $A$, where an ideal of a $C^*$-algebra always means a closed two-sided ideal, except in the proof of Lemma 4.6.
Definition 1.2. For a Hilbert $A$-module $X$, we denote by $\mathcal{L}(X)$ the $C^*$-algebra of all adjointable operators on $X$. For $\xi, \eta \in X$, the operator $\theta_{\xi,\eta} \in \mathcal{L}(X)$ is defined by $\theta_{\xi,\eta}(\zeta) = \xi \langle \eta, \zeta \rangle_X$ for $\zeta \in X$. We define the ideal $\mathfrak{K}(X)$ of $\mathcal{L}(X)$ by

$$\mathfrak{K}(X) = \overline{\text{span}} \{ \theta_{\xi,\eta} \in \mathcal{L}(X) \mid \xi, \eta \in X \}.$$ 

We fix a $C^*$-algebra $A$ and a Hilbert $A$-module $X$ throughout this section.

Proposition 1.3. Let $I$ be an ideal of $A$. For $\xi \in X$, the following are equivalent:

(i) $\xi \in X I$.
(ii) $\langle \eta, \xi \rangle_X \in I$ for all $\eta \in X$.
(iii) $\langle \xi, \xi \rangle_X \in I$.
(iv) There exist $\eta \in X$ and a positive element $a \in I$ such that $\xi = \eta a$.

Proof. Clearly (iv) $\Rightarrow$ (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii). For $\xi \in X$ with $\langle \xi, \xi \rangle_X \in I$, we can find $\eta \in X$ such that $\xi = \eta a$ for $a = (\langle \xi, \xi \rangle_X)^{1/3} \in I$ ([Lance 1995, Lemma 4.4]). This proves (iii) $\Rightarrow$ (iv). □

Corollary 1.4. For an ideal $I$ of $A$, $XI$ is a closed linear subspace of $X$ which is invariant by the right action of $A$ and by the left action of $\mathcal{L}(X)$.

Proof. Since the set of $\xi \in X$ satisfying condition (ii) in Proposition 1.3 is a closed linear space, we see that $XI$ is a closed linear space (this also follows from the Cohen factorization theorem). The rest of the statement is easy to verify. □

By this corollary, $XI$ is a Hilbert $A$-submodule of $X$. We can and will consider $\mathfrak{K}(XI)$ as a subalgebra of $\mathfrak{K}(X)$ by

$$\mathfrak{K}(XI) = \overline{\text{span}} \{ \theta_{\xi,\eta} \in \mathfrak{K}(X) \mid \xi, \eta \in XI \} \subset \mathfrak{K}(X)$$

(see [Fowler et al. 2003, Lemma 2.6 (1)] for the proof). Note that $XI$ is also considered as a Hilbert $I$-module. For an ideal $I$ of $A$, we denote by $X_I$ the quotient space $X/XI$. Both of the natural quotient maps $A \to A/I$ and $X \to X_I$ are denoted by $[\cdot]_I$. The space $X_I$ has an $A/I$-valued inner product $\langle \cdot, \cdot \rangle_{X_I}$ and a right action of $A/I$ so that

$$\langle [\xi]_I, [\zeta]_I \rangle_{X_I} = \langle \xi, \zeta \rangle_X, \quad [\xi]_I[a]_I = [\xi a]_I$$

for $\xi, \zeta \in X$ and $a \in A$. By Proposition 1.3, $\eta \in X_I$ satisfies $\langle \eta, \eta \rangle_{X_I} = 0$ only when $\eta = 0$. Hence $\|\eta\|_{X_I} = \|\langle \eta, \eta \rangle_{X_I}\|^{1/2}$ defines a norm on $X_I$.

Lemma 1.5. For $\eta \in X_I$, there exists $\xi \in X$ such that $\eta = [\xi]_I$ and $\|\eta\|_{X_I} = \|\xi\|_X$. 

Proof: Clearly \([ \cdot ]_I\) is a norm-decreasing map. Thus it suffices to find \(\xi \in X\) such that \([\xi]_I = \eta\) and \(\|\xi\|_X \leq \|\eta\|_{X_I}\) for \(\eta \in X_I\). Set \(C = \|\eta\|^2_{X_I} = \|\langle \eta, \eta \rangle_{X_I}\|\). Let \(f, g\) be functions on \(\mathbb{R}_+ = [0, \infty)\) defined by

\[
    f(r) = \begin{cases} 
        1 & \text{if } 0 \leq r \leq C \\
        \sqrt{C/r} & \text{if } r > C
    \end{cases},
    g(r) = \min\{r, C\}.
\]

Then we have \(g(r) = rf(r)^2\) and \(g(r) \leq C\) for \(r \in \mathbb{R}_+\). Take \(\xi_0 \in X\) with \(\eta = [\xi_0]_I\). Set \(a = f(\langle \xi_0, \xi_0 \rangle_X) \in A\) and \(\xi = \xi_0a \in X\) where \(A\) is the unitalization of \(A\). We have \(\langle \xi, \xi \rangle_X = a^*(\xi_0, \xi_0)a = g(\langle \xi_0, \xi_0 \rangle_X)\). Hence we get \(\|\xi\|_X \leq C^{1/2} = \|\eta\|_{X_I}\).

Since \(f\) is 1 on \([0, C]\), we have

\[
    [a]_I = f(\langle \xi_0, \xi_0 \rangle_X) = f(\langle \eta, \eta \rangle_{X_I}) = 1.
\]

Therefore we see that \([\xi]_I = [\xi_0]_I[a]_I = \eta\). We are done. \(\square\)

By this lemma, the norm \(\| \cdot \|_{X_I}\) of \(X_I\) coincides with the quotient norm of \([\cdot ]_I: X \to X_I\) (see [Fowler et al. 2003, Lemma 2.1] for another proof). Hence \(X_I\) is complete, and so it is a Hilbert \(A/I\)-module.

Since \(XI\) is closed under the action of \(\mathcal{L}(X)\), we can define a map \(\mathcal{L}(X) \to \mathcal{L}(X_I)\), which is also denoted by \([\cdot ]_I\), so that \([S]_I[\xi]_I = [S\xi]_I\) for \(S \in \mathcal{L}(X)\) and \(\xi \in X\). By definition, \(S \in \mathcal{L}(X)\) satisfies \([S]_I = 0\) if and only if \(S\xi \in XI\) for all \(\xi \in X\), which is equivalent to the condition that \(\langle \eta, S\xi \rangle \in I\) for all \(\xi, \eta \in X\) by Proposition 1.3.

**Lemma 1.6.** For \(\xi, \eta \in X\), we have \([\theta_{\xi, \eta}]_I = [\theta_{[\xi]_I, [\eta]_I}]_I\). The restriction of the map \([\cdot ]_I: \mathcal{L}(X) \to \mathcal{L}(X_I)\) to \(\mathcal{K}(X)\) is a surjection onto \(\mathcal{K}(XI)\) whose kernel is \(\mathcal{K}(XI)\).

**Proof.** The first assertion is easily verified by the definition. This implies that the restriction of the map \([\cdot ]_I\) to \(\mathcal{K}(X)\) is a surjection onto \(\mathcal{K}(XI)\), and that \(\mathcal{K}(XI)\) is in the kernel of \([\cdot ]_I\). We will show that if \(k \in \mathcal{K}(X)\) satisfies that \([k]_I = 0\), then \(k \in \mathcal{K}(XI)\).

There exists an approximate unit \((u_\lambda)_{\lambda \in \Lambda}\) of \(\mathcal{K}(X)\) such that for each \(\lambda \in \Lambda\), \(u_\lambda\) is a finite linear sum of elements in the form \(\theta_{\xi, \eta}\). Take \(k \in \mathcal{K}(X)\) with \([k]_I = 0\). Since we have \(k = \lim k\theta_{\xi, \eta}\), to prove \(k \in \mathcal{K}(XI)\) it suffices to show that \(k\theta_{\xi, \eta} \in \mathcal{K}(XI)\) for arbitrary \(\xi, \eta \in X\). Since \(k\xi \in XI\), we can find \(\xi_0 \in X\) and a positive element \(a_0 \in I\) such that \(k\xi = \xi_0a_0\) by Proposition 1.3. Then we have

\[
    k\theta_{\xi, \eta} = \theta_{k\xi, \eta} = \theta_{\xi_0a_0, \eta} = \theta_{\xi_0, \eta} \sqrt{a_0} \sqrt{a_0} \in \mathcal{K}(XI),
\]

as needed. \(\square\)

See also [Fowler et al. 2003, Lemma 2.6 (2), (3)]. Note that it often happens that \([S]_I \in \mathcal{K}(XI)\) even if \(S \notin \mathcal{K}(X)\). This observation plays an important role in our analysis after Section 5. Note also that though three maps \([\cdot ]_I: A \to A/I, [\cdot ]_I: X \to X_I\) and \([\cdot ]_I: \mathcal{K}(X) \to \mathcal{K}(XI)\) are always surjective, the map
[\cdot]_I : \mathcal{L}(X) \to \mathcal{L}(X_I) need not be surjective, because Tietze’s extension theorem fails in general.

Take two ideals $I$ and $I'$ of $A$ such that $I \subset I'$. Then $I'/I$ is an ideal of $A/I$ and $(A/I)/(I'/I) \ni [a]_I \mapsto [a]_{I'} \in A/I'$ gives a well-defined isomorphism. By this isomorphism, we will identify $(A/I)/(I'/I)$ with $A/I'$. Thus the quotient map $[\cdot]_{I'} : A \to A/I'$ coincides with the composition of $[\cdot]_I : A \to A/I$ and $[\cdot]_{I'/I} : A/I \to A/I'$. Similarly we will identify $(X_I)_{I'/I}$ with $X_{I'}$ so that $[\cdot]_{I'} = [\cdot]_{I'/I} \circ [\cdot]_I$ holds for both $X \to X_{I'}$ and $\mathcal{L}(X) \to \mathcal{L}(X_{I'})$. It is easy to see the following.

**Lemma 1.7.** We have $(XI')_1 = X_I(I'/I)$ in $X_I$.

Now take two ideals $I_1$ and $I_2$ of $A$. It is well-known that the ideal $I_1 \cap I_2$ coincides with $I_1I_2$, and that $I_1 + I_2$ is an ideal of $A$. It is easy to see that the natural map $I_1/(I_1 \cap I_2) \to (I_1 + I_2)/I_2$ is an isomorphism. The pull-back $C^*$-algebra $B$ of the two quotient maps $[\cdot]_{I_1+I_2} : A/I_1 \to A/(I_1 + I_2)$ and $[\cdot]_{I_1+I_2} : A/I_2 \to A/(I_1 + I_2)$ is defined by

$$B = \{(b_1, b_2) \in A/I_1 \oplus A/I_2 \mid [b_1]_{I_1+I_2}/I_1 = [b_2]_{I_1+I_2}/I_2 \in A/(I_1 + I_2)\}.$$  

It is not difficult to see the following (see the proof of Proposition 1.10).

**Lemma 1.8.** The map

$$\Pi : A/(I_1 \cap I_2) \ni b \mapsto ([b]_{I_1/(I_1 \cap I_2)}, [b]_{I_2/(I_1 \cap I_2)}) \in B$$

is an isomorphism.

We will show analogous statements for Hilbert modules and sets of operators on them. Define a linear space $Y$ by

$$Y = \{ (\eta_1, \eta_2) \in X_{I_1} \oplus X_{I_2} \mid [\eta_1]_{I_1+I_2}/I_1 = [\eta_2]_{I_1+I_2}/I_2 \in X_{I_1+I_2}\}.$$  

We define a $B$-valued inner product on $Y$ by

$$\langle (\eta_1, \eta_2), (\eta'_1, \eta'_2) \rangle_Y = \langle \eta_1, \eta'_1 \rangle_{X_{I_1}}, \langle \eta_2, \eta'_2 \rangle_{X_{I_2}} \rangle \in B,$$

for $(\eta_1, \eta_2), (\eta'_1, \eta'_2) \in Y$. Clearly $Y$ is complete with respect to the norm defined by the inner product. If we define a right action of $B$ on $Y$ by

$$\eta_1 (b_1, b_2) = \eta_1 b_1, \eta_2 b_2 \in Y$$

for $(\eta_1, \eta_2) \in Y, (b_1, b_2) \in B$, then we can easily see that $Y$ is a Hilbert $B$-module.

**Lemma 1.9.** The restriction of the quotient map $[\cdot]_{I_2/(I_1 \cap I_2)} : X_{I_1 \cap I_2} \to X_{I_2}$ to $X_{I_1 \cap I_2}((I_1 + I_2)/I_2)$ is a bijection onto $X_{I_2}((I_1 + I_2)/I_2)$.
Proof: By Lemma 1.7, we have $X_{I_1 \cap I_2}((I_1 / (I_1 \cap I_2)) = (XI_1)_{I_1 \cap I_2}$. It is easy to see that the surjection $\sigma_{I_2 / (I_1 \cap I_2)} : (XI_1)_{I_1 \cap I_2} \to (X I_1)_{I_2}$ is injective. It is also easy to see that $(X I_1)_{I_2} = (X (I_1 + I_2))_{I_2}$. We have $(X (I_1 + I_2))_{I_2} = X_{I_2}((I_1 + I_2)/I_2)$ by Lemma 1.7. This completes the proof.

Proposition 1.10. By $\Pi$ in Lemma 1.8, we can consider $X_{I_1 \cap I_2}$ as a Hilbert $B$-module. Then the map

$$T : X_{I_1 \cap I_2} \ni \eta \mapsto ([\eta]_{I_1 / (I_1 \cap I_2)}, [\eta]_{I_2 / (I_1 \cap I_2)}) \in Y$$

is an isomorphism of Hilbert $B$-modules.

Proof. Clearly $T$ preserves inner products and right actions. This implies that $T$ is isometric. It remains to show that $T$ is surjective. Take $(\eta_1, \eta_2) \in Y$. Since $\cdot \cdot_{I_1 / (I_1 \cap I_2)} : X_{I_1 \cap I_2} \to X_{I_1}$ is surjective, we can find $\eta' \in X_{I_1 \cap I_2}$ with $[\eta']_{I_1 / (I_1 \cap I_2)} = \eta_1$. Since $[\eta_2]_{I_1 + I_2 / I_2} = [\eta_1]_{I_1 + I_2 / I_2}$, we have

$$\eta_2 - [\eta']_{I_2 / (I_1 \cap I_2)} \in \ker ([\cdot]_{I_1 + I_2 / I_2}) = X_{I_2}((I_1 + I_2)/I_2).$$

By Lemma 1.9, we can find $\eta'' \in X_{I_1 \cap I_2}(I_1 / (I_1 \cap I_2))$ with

$$[\eta'']_{I_2 / (I_1 \cap I_2)} = \eta_2 - [\eta']_{I_2 / (I_1 \cap I_2)},$$

Set $\eta = \eta' + \eta'' \in X_{I_1 \cap I_2}$. We see that

$$[\eta]_{I_1 / (I_1 \cap I_2)} = [\eta']_{I_1 / (I_1 \cap I_2)} + 0 = \eta_1,$n

$$[\eta]_{I_2 / (I_1 \cap I_2)} = [\eta']_{I_2 / (I_1 \cap I_2)} + [\eta'']_{I_2 / (I_1 \cap I_2)} = \eta_2.$$n

Therefore $T(\eta) = (\eta_1, \eta_2).$ Thus $T$ is surjective. \qed

Proposition 1.11. Define a $C^*$-algebra $\mathcal{M}$ by

$$\mathcal{M} = \{ (S_1, S_2) \in \mathcal{L}(X_{I_1}) \oplus \mathcal{L}(X_{I_2}) \mid [S_1]_{I_1 + I_2 / I_1} = [S_2]_{I_1 + I_2 / I_2} \in \mathcal{L}(X_{I_1 + I_2}) \}.$$n

Then the map

$$\Psi : \mathcal{L}(X_{I_1 \cap I_2}) \ni S \mapsto ([S]_{I_1 / (I_1 \cap I_2)}, [S]_{I_2 / (I_1 \cap I_2)}) \in \mathcal{M}$$

is an isomorphism, and its restriction to $\mathcal{K}(X_{I_1 \cap I_2})$ is an isomorphism onto the $C^*$-subalgebra $\mathcal{K}$ of $\mathcal{M}$ defined by

$$\mathcal{K} = \{ (k_1, k_2) \in \mathcal{K}(X_{I_1}) \oplus \mathcal{K}(X_{I_2}) \mid [k_1]_{I_1 + I_2 / I_1} = [k_2]_{I_1 + I_2 / I_2} \in \mathcal{K}(X_{I_1 + I_2}) \}.$$n

Proof. Take $(S_1, S_2) \in \mathcal{M}$ and define $\Psi'(S_1, S_2) \in \mathcal{L}(X_{I_1 \cap I_2})$. For $\xi \in X_{I_1 \cap I_2}$, we have

$$[S_2[\xi]_{I_1 / (I_1 \cap I_2)}]_{I_1 + I_2 / I_1} = [S_2[\xi]_{I_2 / (I_1 \cap I_2)}]_{I_1 + I_2 / I_2}.$$

Hence by Proposition 1.10, there exists a unique element $\eta \in X_{I_1 \cap I_2}$ with

$$[\eta]_{I_1 / (I_1 \cap I_2)} = S_1[\xi]_{I_1 / (I_1 \cap I_2)}, \quad \text{and} \quad [\eta]_{I_2 / (I_1 \cap I_2)} = S_2[\xi]_{I_2 / (I_1 \cap I_2)}.\]n
We define $\Psi'(S_1, S_2) : X_{I_1 \cap I_2} \to X_{I_1 \cap I_2}$ by $\Psi'(S_1, S_2)\xi = \eta$ where $\eta$ is the unique element satisfying the two equations above. Then, using Lemma 1.8, we see that
\[
\langle \Psi'(S_1, S_2)\xi, \xi' \rangle_{X_{I_1 \cap I_2}} = \langle \xi, \Psi'(S_1^*, S_2^*)\xi' \rangle_{X_{I_1 \cap I_2}}
\]
for every $\xi, \xi' \in X_{I_1 \cap I_2}$. Thus $\Psi'(S_1, S_2) \in \mathcal{L}(X_{I_1 \cap I_2})$ for all $(S_1, S_2) \in \mathcal{M}$. It is easy to see that $\Psi' : \mathcal{M} \to \mathcal{L}(X_{I_1 \cap I_2})$ is a *-homomorphism, and gives the inverse of $\Psi$. Hence $\Psi : \mathcal{L}(X_{I_1 \cap I_2}) \to \mathcal{M}$ is an isomorphism.

Clearly the restriction of $\Psi$ on $\mathcal{H}(X_{I_1 \cap I_2})$ is an injection into $\mathcal{H}$. We will show that this is surjective. By Lemma 1.9, we can see that the restriction of the map $[\cdot]_{I_2/(I_1 \cap I_2)} : \mathcal{H}(X_{I_1 \cap I_2}) \to \mathcal{H}(X_{I_2})$ to
\[
\ker([\cdot]_{I_1/(I_1 \cap I_2)}) = \mathcal{H}(X_{I_1 \cap I_2}(I_1/(I_1 \cap I_2)))
\]
is a bijection onto
\[
\ker([\cdot]_{I_1+I_2/I_2}) = \mathcal{H}(X_{I_2}((I_1+I_2)/I_2)).
\]
Take $(k_1, k_2) \in \mathcal{H}$. Since the map $[\cdot]_{I_1/(I_1 \cap I_2)} : \mathcal{H}(X_{I_1 \cap I_2}) \to \mathcal{H}(X_{I_1})$ is surjective, we can find $k' \in \mathcal{H}(X_{I_1 \cap I_2})$ with $[k']_{I_1/(I_1 \cap I_2)} = k_1$. Then we see that
\[
k_2 - [k']_{I_2/(I_1 \cap I_2)} \in \ker([\cdot]_{I_1+I_2/I_2}).
\]
Thus there exists a unique element
\[
k'' \in \ker([\cdot]_{I_1/(I_1 \cap I_2)}) \subset \mathcal{H}(X_{I_1 \cap I_2})
\]
with $[k'']_{I_2/(I_1 \cap I_2)} = k_2 - [k']_{I_2/(I_1 \cap I_2)}$. Now it is easy to see that $k = k' + k'' \in \mathcal{H}(X_{I_1 \cap I_2})$ satisfies $\Psi(k) = (k_1, k_2)$. We are done. \hfill \Box

**Corollary 1.12.** If $S \in \mathcal{L}(X_{I_1 \cap I_2})$ satisfies
\[
[S]_{I_1/(I_1 \cap I_2)} \in \mathcal{H}(X_{I_1}), \quad [S]_{I_2/(I_1 \cap I_2)} \in \mathcal{H}(X_{I_2}),
\]
then $S \in \mathcal{H}(X_{I_1 \cap I_2})$.

**Proof:** Clear by Proposition 1.11. \hfill \Box

2. $C^*$-correspondences and representations

**Definition 2.1.** For a $C^*$-algebra $A$, we say that $X$ is a $C^*$-correspondence over $A$ when $X$ is a Hilbert $A$-module and a *-homomorphism $\varphi_X : A \to \mathcal{L}(X)$ is given.

We refer to $\varphi_X$ as the left action of a $C^*$-correspondence $X$. $C^*$-correspondences can be considered as generalizations of automorphisms or endomorphisms. In fact, we can associate a $C^*$-correspondence $X_\varphi$ with each endomorphism $\varphi$ as follows.
Definition 2.2. Let $A$ be a $C^*$-algebra and $\varphi: A \to A$ be an endomorphism. We define a $C^*$-correspondence $X_\varphi$ such that it is isomorphic to $A$ as Banach spaces, its inner product is defined by $\langle \xi, \eta \rangle_X = \xi^* \eta$, right action is multiplication and left action is given by $\varphi_{X_\varphi}(a)\xi = \varphi(a)\xi$. We denote $X_{\text{id}_A}$ by $A$, and call it the identity correspondence over $A$.

Note that the left action $\varphi_A$ of the identity correspondence $A$ gives an isomorphism from $A$ to $\mathcal{K}(A)$.

Definition 2.3. A morphism from a $C^*$-correspondence $X$ over a $C^*$-algebra $A$ to a $C^*$-correspondence $Y$ over a $C^*$-algebra $B$ is a pair $(\Pi, T)$ consisting of a $*$-homomorphism $\Pi: A \to B$ and a linear map $T: X \to Y$ satisfying

(i) $\{T(\xi), T(\eta)\}_Y = \Pi(\{\xi, \eta\}_X)$ for $\xi, \eta \in X$,
(ii) $\varphi_Y(\Pi(a))T(\xi) = T(\varphi_X(a)\xi)$ for $a \in A, \xi \in X$.

A morphism $(\Pi, T)$ is said to be injective if a $*$-homomorphism $\Pi$ is injective.

A morphism is called a semicovariant homomorphism in [Schweizer 2001]. For a morphism $(\Pi, T)$ from $X$ to $Y$, we can see that $T(\xi)\Pi(a) = T(\xi) = \xi$ for $a \in A$ and $\xi \in X$ by the same argument as in [Katsura 2004b, Section 2]. We also see that $T$ is isometric for an injective morphism $(\Pi, T)$.

Definition 2.4. For a morphism $(\Pi, T)$ from a $C^*$-correspondence $X$ over $A$ to a $C^*$-correspondence $Y$ over $B$, we define a $*$-homomorphism $\Psi_T: \mathcal{K}(X) \to \mathcal{K}(Y)$ by $\Psi_T(\theta_{\xi, \eta}) = \theta_{T(\xi), T(\eta)}$ for $\xi, \eta \in X$.

For the well-definedness of a $*$-homomorphism $\Psi_T$, see, for example, [Kajiwara et al. 1998, Lemma 2.2]. Note that $\Psi_T$ is injective for an injective morphism $(\Pi, T)$. The following two lemmas are easily verified.

Lemma 2.5. For a morphism $(\Pi, T)$ from a $C^*$-correspondence $X$ over $A$ to a $C^*$-correspondence $Y$ over $B$, we have $\varphi_Y(\Pi(a))\Psi_T(\xi) = \Psi_T(\varphi_X(a)k)$ and $\Psi_T(k)T(\xi) = T(k\xi)$ for $a \in A, \xi \in X$ and $k \in \mathcal{K}(X)$.

Lemma 2.6. Let $X, Y, Z$ be $C^*$-correspondences, and $(\Pi_1, T_1), (\Pi_2, T_2)$ be morphisms from $X$ to $Y$ and from $Y$ to $Z$, respectively. Then its composition $(\Pi_2 \circ \Pi_1, T_2 \circ T_1)$ is a morphism from $X$ to $Z$, and we have $\Psi_{T_2 \circ T_1} = \Psi_{T_2} \circ \Psi_{T_1}$.

Definition 2.7. A representation of a $C^*$-correspondence $X$ over $A$ on a $C^*$-algebra $B$ is a pair $(\pi, t)$ consisting of a $*$-homomorphism $\pi: A \to B$ and a linear map $t: X \to B$ satisfying

(i) $t(\xi^*)t(\eta) = \pi(\{\xi, \eta\}_X)$ for $\xi, \eta \in X$,
(ii) $\pi(a)t(\xi) = t(\varphi_X(a)\xi)$ for $a \in A, \xi \in X$.

We denote by $C^*(\pi, t)$ the $C^*$-algebra generated by the images of $\pi$ and $t$ in $B$. We define a $*$-homomorphism $\psi_t: \mathcal{K}(X) \to C^*(\pi, t)$ by $\psi_t(\theta_{\xi, \eta}) = t(\xi)t(\eta)^* \in C^*(\pi, t)$ for $\xi, \eta \in X$. 
Representations of a C*-correspondence $X$ on a C*-algebra $B$ are precisely the morphisms from $X$ to the identity correspondence over $B$, and we have $\varphi_B \circ \psi_t = \Psi_t$. Note that we get $\pi(a) \psi_t(k) = \psi_t(\varphi_X(a)k)$ and $\psi_t(k) t(\xi) = t(k\xi)$ for $k \in \mathcal{H}(X)$, $a \in A$ and $\xi \in X$.

**Definition 2.8.** A representation $(\pi, t)$ of $X$ is said to admit a gauge action if for each $z \in \mathbb{T}$, there exists a $\ast$-homomorphism $\beta_z : C^*(\pi, t) \to C^*(\pi, t)$ such that $\beta_z(\pi(a)) = \pi(a)$ and $\beta_z(t(\xi)) = zt(\xi)$ for all $a \in A$ and $\xi \in X$.

If it exists, such a $\ast$-homomorphism $\beta_z$ is unique and $\beta : \mathbb{T} \to \text{Aut}(C^*(\pi, t))$ is a strongly continuous homomorphism.

### 3. C*-algebras associated with C*-correspondences

In this section, we review the constructions of the C*-algebras $\mathcal{F}_X$ and $\mathcal{O}_X$ from a C*-correspondence $X$. These C*-algebras were introduced by Pimsner in [Pimsner 1997], and modified in [Katsura 2003a].

**Definition 3.1.** For a C*-correspondence $X$ over a C*-algebra $A$, we denote by $\mathcal{F}_X$ the C*-algebra generated by the universal representation.

The universal representation can be obtained by taking a direct sum of sufficiently many representations. By universality, we have a surjection $\mathcal{F}_X \to C^*(\pi, t)$ for every representation $(\pi, t)$ of $X$. The C*-algebra $\mathcal{F}_X$ is too large to reflect the informations of $X$, and so we will take a certain quotient of $\mathcal{F}_X$ to get the nice C*-algebra $\mathcal{O}_X$.

**Definition 3.2.** For an ideal $I$ of a C*-algebra $A$, we define $I^\perp \subset A$ by

$$I^\perp = \{a \in A \mid ab = 0 \text{ for all } b \in I\}.$$  

Note that $I^\perp$ is the largest ideal of $A$ satisfying $I \cap I^\perp = 0$.

**Definition 3.3.** For a C*-correspondence $X$ over $A$, we define an ideal $J_X$ of $A$ by

$$J_X = \varphi_X^{-1}\left(\mathcal{H}(X)\right) \cap \left(\ker \varphi_X\right)^\perp.$$  

The ideal $J_X$ is the largest ideal to which the restriction of $\varphi_X$ is an injection into $\mathcal{H}(X)$.

**Definition 3.4.** A representation $(\pi, t)$ of $X$ is said to be covariant if we have $\pi(a) = \psi_t(\varphi_X(a))$ for all $a \in J_X$.

**Definition 3.5.** For a C*-correspondence $X$ over a C*-algebra $A$, the C*-algebra $\mathcal{O}_X$ is defined by $\mathcal{O}_X = C^*(\pi_X, t_X)$ where $(\pi_X, t_X)$ is the universal covariant representation of $X$. 

By universality, for any covariant representation \((\pi, t)\) of a \(C^*\)-correspondence \(X\), there exists a \(*\)-homomorphism \(\rho_{(\pi, t)}: \mathcal{O}_X \to C^* (\pi, t)\) such that \(\pi = \rho_{(\pi, t)} \circ \pi_X\) and \(t = \rho_{(\pi, t)} \circ t_X\). Again by universality, the universal covariant representation \((\pi_X, t_X)\) admits a gauge action. We denote it by \(\gamma: \mathbb{T} \curvearrowright \mathcal{O}_X\). When we consider \(\mathcal{O}_X\) as a generalization of crossed products by automorphisms, the gauge action \(\gamma\) is regarded as the dual action of \(\mathbb{T}\). If a covariant representation \((\pi, t)\) admits a gauge action \(\beta\), then we have \(\beta_z \circ \rho_{(\pi, t)} = \rho_{(\pi, t)} \circ \gamma_z\) for each \(z \in \mathbb{T}\). In [Katsura 2004b, Proposition 4.11], we saw that the universal covariant representation \((\pi_X, t_X)\) is injective. The following gauge-invariant uniqueness theorem says that two conditions, admitting a gauge action and being injective, characterize the universal one \((\pi_X, t_X)\) among all covariant representations.

**Theorem 3.6** [Katsura 2004b, Theorem 6.4]. For a covariant representation \((\pi, t)\) of a \(C^*\)-correspondence \(X\), the map \(\rho_{(\pi, t)}: \mathcal{O}_X \to C^* (\pi, t)\) is an isomorphism if and only if \((\pi, t)\) is injective and admits a gauge action.

In Proposition 7.14, we see that the universal covariant representation \((\pi_X, t_X)\) is the smallest one among injective representations admitting gauge actions.

**Remark 3.7.** A morphism \((\Pi, T)\) from a \(C^*\)-correspondence \(X\) to a \(C^*\)-correspondence \(Y\) gives us a \(*\)-homomorphism \(\mathcal{F}_X \to \mathcal{F}_Y\). This also gives a \(*\)-homomorphism \(\mathcal{O}_X \to \mathcal{O}_Y\) when the morphism \((\Pi, T)\) is covariant, that is, we have \(\Pi(a) \in \mathcal{O}_Y\) and \(\varphi_Y (\Pi(a)) = \Psi_T(\varphi_X(a))\) for all \(a \in \mathcal{O}_X\). We do not use these facts explicitly.

### 4. Invariant ideals

In this section, we introduce the notion of invariant ideals with respect to \(C^*\)-correspondences. Let us take a \(C^*\)-correspondence \(X\) over a \(C^*\)-algebra \(A\), and fix them until the end of Section 9.

**Definition 4.1.** For an ideal \(I\) of \(A\), we define \(X(I), X^{-1}(I) \subset A\) by

\[
X(I) = \overline{\text{span}} \left\{ \langle \eta, \varphi_X(a)\xi \rangle \in A \mid a \in I, \xi, \eta \in X \right\},
\]

\[
X^{-1}(I) = \{ a \in A \mid \langle \eta, \varphi_X(a)\xi \rangle_X \in I \text{ for all } \xi, \eta \in X \}.
\]

Clearly \(X(I)\) is an ideal of \(A\). We also see that \(X^{-1}(I)\) is an ideal because it is the kernel of the composition of \(\varphi_X\) and the map \([\cdot]_I: \mathcal{L}(X) \to \mathcal{L}(X_I)\). For a \(C^*\)-correspondence \(X_\varphi\) defined from an endomorphism \(\varphi: A \to A\), we see that \(X_\varphi(I)\) is the ideal generated by \(\varphi(I)\), and \(X_{\varphi}^{-1}(I) = \varphi^{-1}(I)\) for an ideal \(I\) of \(A\). It is easy to see that we have \(X(I_1) \subset X(I_2)\) and \(X^{-1}(I_1) \subset X^{-1}(I_2)\) for two ideals \(I_1, I_2\) of \(A\) with \(I_1 \subset I_2\). For an ideal \(I\), we have \(X(X^{-1}(I)) \subset I\) and \(X^{-1}(X(I)) \supset I\). These inclusions are proper in general, because we always have \(X(I) \subset \overline{\text{span}}(X, X)_X\) and
For two ideals $I$, $I$ let $\phi: A \to A$ be an endomorphism defined by $\phi(\lambda, \mu, T) = (0, 0, \text{diag}\{\lambda, \mu\})$. This endomorphism gives us a $C^*$-correspondence $X = X_\phi$ over $A$. Let us define three ideals $I_1$, $I_2$ and $I_3$ of $A$ by $I_1 = C \oplus 0 \oplus 0$, $I_2 = 0 \oplus C \oplus 0$ and $I_3 = 0 \oplus 0 \oplus M_2(\mathbb{C})$. We see that $\ker \phi_X = \ker \phi = I_3$ and $\phi^{-1}_X(\mathfrak{L}(X)) = A$. Hence we get $J_X = I_1 + I_2$. We have $X(I_1) = X(I_2) = I_3$. However clearly we have $X(I_1 \cap I_2) = X(0) = 0$. This gives an example of a proper inclusion $X(I_1 \cap I_2) \subset X(I_1) \cap X(I_2)$. Since $X^{-1}(I_3) = A$, we have two proper inclusions $X^{-1}(X(I_i)) \supset I_i + \ker \phi_X$ for $i = 1, 2$. We see that there exist no nontrivial invariant ideals of $A$ (see Definition 4.8), and the $C^*$-algebra $\mathfrak{O}_X$ is isomorphic to a simple $C^*$-algebra $M_6(\mathbb{C})$.

For an increasing family $\{I_n\}_{n \in \mathbb{N}}$ of ideals of a $C^*$-algebra $D$, we denote by $\lim_{n \to \infty} I_n$ the ideal of $D$ defined by

$$\lim_{n \to \infty} I_n = \bigcup_{n \in \mathbb{N}} I_n.$$

**Proposition 4.4.** Let $\{I_n\}_{n \in \mathbb{N}}$ be an increasing family of ideals of $A$. Then we have $X(\lim_{n \to \infty} I_n) = \lim_{n \to \infty} X(I_n)$.

**Proof.** Clear by the definition of $X(\cdot)$.

The analogous statement of Proposition 4.4 for $X^{-1}$ is not valid as the next example shows.

**Example 4.5.** Let $A = C((0, 1])$. We define a $C^*$-correspondence $X$ over $A$ which is isomorphic to $A$ as Hilbert $A$-modules and its left action $\varphi_X: A \to \mathfrak{L}(X)$ is defined by $\varphi_X(f) = f(1) \text{id}_X$ for $f \in A$. For each $n \in \mathbb{N}$, we define an ideal $I_n$ of $A$ by $I_n = C((2^{-n}, 1])$. We have $\lim_{n \to \infty} I_n = 0$. It is not difficult to see that $X^{-1}(I_n) = C((0, 1))$ for every $n \in \mathbb{N}$. Hence we get $\lim_{n \to \infty} X^{-1}(I_n) = C((0, 1))$. However, we have $X^{-1}(\lim_{n \to \infty} I_n) = X^{-1}(A) = A$. The $C^*$-algebra
\(C_X\) is isomorphic to the universal C*-algebra generated by a contractive scaling element (see [Katsura 2006b]).

Though we do not have \(X^{-1}(\lim_{n\to\infty} I_n) = \lim_{n\to\infty} X^{-1}(I_n)\) in general, we can prove Proposition 4.7, which suffices for the further investigation. For the proof of Proposition 4.7, we need the following general fact.

**Lemma 4.6.** Let \(D\) be a C*-algebra, and \(\{I_n\}_{n\in\mathbb{N}}\) be an increasing family of ideals of \(D\). For each C*-subalgebra \(B\) of \(D\), we have \(B(\lim_{n\to\infty} I_n) = \lim_{n\to\infty}(B \cap I_n)\).

**Proof.** Set \(I_\infty = \lim_{n\to\infty} I_n\). Clearly we have \(B \cap I_\infty = \lim_{n\to\infty}(B \cap I_n)\). Take a positive element \(x \in B \cap I_\infty\). For \(\varepsilon > 0\), let \(f_\varepsilon : \mathbb{R}_+ \to \mathbb{R}_+\) be a continuous function defined by \(f_\varepsilon(t) = \max\{0, t-\varepsilon\}\). Then we have \(\|x - f_\varepsilon(x)\| \leq \varepsilon\). Since \(\bigcup_{n\in\mathbb{N}} I_n\) is a dense ideal in \(I_\infty\), we have \(f_\varepsilon(x) \in \bigcup_{n\in\mathbb{N}} I_n\) (see [Pedersen 1979, Theorem 5.6.1]).

Thus \(x\) is approximated by elements \(f_\varepsilon(x) \in B \cap \bigcup_{n\in\mathbb{N}} I_n = \bigcup_{n\in\mathbb{N}} (B \cap I_n)\). This shows that \(x \in \lim_{n\to\infty}(B \cap I_n)\). Therefore \(B(\lim_{n\to\infty} I_n) = \lim_{n\to\infty}(B \cap I_n)\). \(\square\)

Note that Lemma 4.6 is not valid when \(I_n\)'s are just C*-subalgebras.

**Proposition 4.7.** Let \(\{I_n\}_{n\in\mathbb{N}}\) be an increasing family of ideals of \(A\). For each ideal \(J\) of \(A\) with \(\varphi_X(J) \subset \mathcal{H}(X)\), we have \(J \cap X^{-1}(\lim_{n\to\infty} I_n) = \lim_{n\to\infty}(J \cap X^{-1}(I_n))\).

**Proof.** Set \(I_\infty = \lim_{n\to\infty} I_n\). First note that we have

\[ J \cap X^{-1}(I) = \{ a \in J \mid \varphi_X(a) \in \mathcal{H}(X) \} \]

for an ideal \(I\) of \(A\) by Lemma 1.6. Take \(a \in J \cap X^{-1}(I_\infty)\) and \(\varepsilon > 0\). It is easy to see that \(\mathcal{H}(XI_\infty) = \lim_{n\to\infty} \mathcal{H}(XI_n)\). By Lemma 4.6, we have \(\varphi_X(J) \cap \mathcal{H}(XI_\infty) = \lim_{n\to\infty}(\varphi_X(J) \cap \mathcal{H}(XI_n))\). Since \(\varphi_X(a) \in \varphi_X(J) \cap \mathcal{H}(XI_\infty)\), we can find \(n \in \mathbb{N}\) and \(k \in \varphi_X(J) \cap \mathcal{H}(XI_n)\) such that \(\|\varphi_X(a) - k\| < \varepsilon\). Then we can find \(x \in J\) with \(\|x\| < \varepsilon\) and \(\varphi_X(x) = \varphi_X(a) - k\). Set \(j = a - x \in J\). We have \(\varphi_X(j) = k \in \mathcal{H}(XI_n)\). Thus we get \(j \in J \cap X^{-1}(I_n)\) and \(\|a - j\| < \varepsilon\). Therefore we get \(J \cap X^{-1}(I_\infty) \subset \lim_{n\to\infty}(J \cap X^{-1}(I_n))\). The converse inclusion is obvious. \(\square\)

**Definition 4.8.** An ideal \(I\) of \(A\) is said to be positively invariant if \(X(I) \subset I\), negatively invariant if \(J_X \cap X^{-1}(I) \subset I\), and invariant if \(I\) is both positively and negatively invariant.

In [Kajiwara et al. 1998; Fowler et al. 2003; Schweizer 2001], a positively invariant ideal is called X-invariant. It is clear that \(I\) is positively invariant if and only if \(I \subset X^{-1}(I)\). It is also equivalent to \(\varphi_X(I)X \subset XI\) by Proposition 1.3. Clearly \(A\) is an invariant ideal. We also see that 0 is invariant because \(X(0) = 0\) and \(J_X \cap X^{-1}(0) = J_X \cap \ker \varphi_X = 0\).

**Proposition 4.9.** Let \(\{I_n\}_{n\in\mathbb{N}}\) be an increasing family of ideals of \(A\). If \(I_n\) is positively invariant (negatively invariant, invariant), then \(\lim_{n\to\infty} I_n\) is also.

**Proposition 4.10.** If two ideals $I_1$, $I_2$ are positively invariant, then their intersection $I_1 \cap I_2$ is also positively invariant. The same is true for negative invariance.

*Proof.* Clear by Lemma 4.2. □

**Corollary 4.11.** The intersection of two invariant ideals is invariant.

By Lemma 4.2, we see that if two ideals $I_1$, $I_2$ are positively invariant, then so is their sum $I_1 + I_2$. However, the sum of two negatively invariant ideals need not be negatively invariant. Moreover, the sum of two invariant ideals can fail to be negatively invariant as we will see in the next example.

**Example 4.12.** Let $A$ be $\mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C}$, and $X$ be $\mathbb{C} \oplus \mathbb{C}$ which is a Hilbert $A$-module by the operations $(\langle \xi_1, \eta_1 \rangle, (\xi_2, \eta_2), \chi) = (\xi_1 \chi \xi_2^*, \eta_1 \eta_2^*)$, and $(\xi, \eta)(\lambda, \mu, \nu) = (\xi \lambda, \eta \nu)$. We define a left action $X$ of $A$ on $\mathcal{L}(X)$ by $\phi_X((\lambda, \mu, \nu)) = \nu \text{id}_X$. We define three ideals $I_1$, $I_2$ and $I_3$ of $A$ by $I_1 = \mathbb{C} \oplus 0 \oplus 0$, $I_2 = 0 \oplus \mathbb{C} \oplus 0$ and $I_3 = 0 \oplus 0 \oplus \mathbb{C}$.

We have $J_X = I_3$. An easy computation shows that $X(I_1) = X(I_2) = 0$ and $X^{-1}(I_1) = X^{-1}(I_2) = I_1 + I_2$. Thus both $I_1$ and $I_2$ are invariant ideals. However we have $X(I_1 + I_2) = 0$ and $X^{-1}(I_1 + I_2) = A$. Thus $I_1 + I_2$ is positively invariant, but not negatively invariant. We also have proper inclusions

$$A = X^{-1}(I_1 + I_2) \supset X^{-1}(I_1) + X^{-1}(I_2) = I_1 + I_2,$$

$$0 = X(X^{-1}(I_i)) \subset I_i \cap \text{span}(X, X) = I_i \quad (i = 1, 2).$$

We have $\mathcal{O}_X \cong M_2(\mathbb{C}) \oplus M_2(\mathbb{C})$, and the two nontrivial invariant ideals $I_1$, $I_2$ correspond to the two nontrivial ideals of $\mathcal{O}_X$.

**Definition 4.13.** Let $A$ be an ideal $I$ of $A$. We define ideals $X^n(I)$ for $n \in \mathbb{N}$ by $X^0(I) = I$ and $X^{n+1}(I) = X(X^n(I))$. We also define ideals $X_{-n}(I)$ for $n \in \mathbb{N}$ by $X_0(I) = I$, $X_{-1}(I) = I + J_X \cap X^{-1}(I)$ and $X_{-(n+1)}(I) = X_{-1}(X_{-n}(I))$ for $n \geq 1$.

Note that we have $I \subset X_{-1}(I)$, hence $X_{-n}(I) \subset X_{-(n+1)}(I)$ for every $n \in \mathbb{N}$.

**Definition 4.14.** For an ideal $I$ of $A$, we define ideals $X^\infty(I)$, $X_{-\infty}(I)$ and $X^{-\infty}(I)$ of $A$ by

$$X^\infty(I) = \sum_{n=0}^{\infty} X^n(I) = \lim_{k \to \infty} \sum_{n=0}^{k} X^n(I), \quad X_{-\infty}(I) = \lim_{n \to \infty} X_{-n}(I),$$

and $X^{-\infty}(I) = X_{-\infty}(X^\infty(I))$.

**Lemma 4.15.** If an ideal $I$ is positively invariant, so are $X_{-n}(I)$ for $n \in \mathbb{N} \cup \{\infty\}$.

*Proof.* Let us take a positively invariant ideal $I$. From

$$X_{-1}(I) = I + J_X \cap X^{-1}(I) \subset X^{-1}(I) \subset X^{-1}(X_{-1}(I))$$
we see that $X_{-1}(I)$ is positively invariant. By using this fact, we can prove inductively that $X_{-n}(I)$ is positively invariant for all $n \in \mathbb{N}$. Finally $X_{-\infty}(I)$ is positively invariant by Proposition 4.9. □

**Proposition 4.16.** For an ideal $I$ of $A$, the ideal $X^{\infty}(I)$ ($X_{-\infty}(I)$, $X^{\infty}_{-\infty}(I)$) is the smallest positively invariant (negatively invariant, invariant) ideal containing $I$.

*Proof.* For each $k \in \mathbb{N}$, we have

$$X\left(\sum_{n=0}^{k} X^n(I)\right) = \sum_{n=0}^{k} X^{n+1}(I) \subset X^{\infty}(I).$$

Hence by Proposition 4.4, we have $X(X^{\infty}(I)) \subset X^{\infty}(I)$. Thus $X^{\infty}(I)$ is positively invariant. If $I'$ is a positively invariant ideal containing $I$, then we can prove inductively $X^n(I) \subset I'$ for all $n \in \mathbb{N}$. Hence we have $X^{\infty}(I) \subset I'$. Thus $X^{\infty}(I)$ is the smallest positively invariant ideal containing $I$.

For each $n \in \mathbb{N}$, we have $J_X \cap X^{-1}(X_{-n}(I)) \subset X_{-(n+1)}(I) \subset X_{-\infty}(I)$. Hence by Proposition 4.7, we have $J_X \cap X^{-1}(X_{-\infty}(I)) \subset X_{-\infty}(I)$. Thus $X_{-\infty}(I)$ is negatively invariant. If $I'$ is a negatively invariant ideal containing $I$, then we can prove inductively $X_{-n}(I) \subset I'$ for all $n \in \mathbb{N}$. Hence we have $X_{-\infty}(I) \subset I'$. Thus $X_{-\infty}(I)$ is the smallest negatively invariant ideal containing $I$.

Combining the above argument with Lemma 4.15, we see that $X^{\infty}_{-\infty}(I)$ is the smallest invariant ideal containing $I$. □

5. **$T$-pairs and $O$-pairs**

In this section, we introduce the notion of $T$-pairs and $O$-pairs of the $C^*$- correspondence $X$ over $A$. These are related to representations of $X$.

**Definition 5.1.** For an ideal $I$ of $A$, we define an ideal $J(I)$ of $A$ by

$$J(I) = \{a \in A \mid [\varphi_X(a)]_I \in \mathcal{L}(X_I), aX^{-1}(I) \subset I\}.$$

For a positively invariant ideal $I$, we can define a map $\varphi_{X_I}: A/I \to \mathcal{L}(X_I)$ so that $\varphi_{X_I}([a]_I) = [\varphi_X(a)]_I$, because $a \in I$ implies $[\varphi_X(a)]_I = 0$. Thus in this case, $X_I$ is a $C^*$-correspondence over $A/I$. It is clear that the pair $(\cdot|_I, [\cdot]_I)$ of the quotient maps $A \to A/I$ and $X \to X_I$ is a morphism from $X$ to $X_I$.

**Lemma 5.2.** For a positively invariant ideal $I$, we have $X^{-1}(I) = [\cdot]_I^{-1}(\ker \varphi_{X_I})$, $J(I) = [\cdot]_I^{-1}(J_X)$, and $X^{-1}(I) \cap J(I) = I$.

*Proof.* We have

$$X^{-1}(I) = \ker([\cdot]|_I \circ \varphi_X) = \ker(\varphi_{X_I} \circ [\cdot]_I) = [\cdot]_I^{-1}(\ker \varphi_{X_I}).$$
We also see that $[\varphi_X(a)]_I \in \mathcal{H}(X_I)$ if and only if $\varphi_X((a)_I) \in \mathcal{H}(X_I)$. Since $X^{-1}(I) = [\cdot]^{-1}_I(\ker \varphi_{X_I})$, the condition $aX^{-1}(I) \subset I$ for $a \in A$ is equivalent to $[a]_I \ker \varphi_{X_I} = 0$. Hence $a \in J(I)$ if and only if

$$[a]_I \in \varphi_{X_I}^{-1}(\mathcal{H}(X_I)) \cap (\ker \varphi_{X_I})^\perp = J_{X_I}.$$ 

Thus we get $J(I) = [\cdot]^{-1}_I(J_{X_I})$. Finally,

$$X^{-1}(I) \cap J(I) = [\cdot]^{-1}_I(\ker \varphi_{X_I} \cap J_{X_I}) = [\cdot]^{-1}_I(0) = I. \quad \square$$

Note that Lemma 5.2 implies that $X^{-1}(I)/I = \ker \varphi_{X_I}$ and $J(I)/I = J_{X_I}$ for a positively invariant ideal $I$. Note also that $X^{-1}(0) = \ker \varphi_X$ and $J(0) = J_X$.

**Proposition 5.3.** An ideal $I$ is negatively invariant if and only if $J_X \subset J(I)$.

**Proof.** For $a \in J_X$, we have $\varphi_X(a) \in \mathcal{H}(X)$. Hence $[\varphi_X(a)]_I \in \mathcal{H}(X_I)$. Thus $J_X \subset J(I)$ if and only if $J_X X^{-1}(I) \subset I$. This is equivalent to the negative invariance of $I$ because $J_X X^{-1}(I) = J_X \cap X^{-1}(I)$. \qed

Note that $I_1 \subset I_2$ need not imply $J(I_1) \subset J(I_2)$ in general as the following example shows.

**Example 5.4** (compare Example 4.12). Let $A \cong \mathbb{C}^3$ be the $C^*$-algebra generated by three mutually orthogonal projections $p_0$, $p_1$ and $p_2$. Let $X$ be the $\ell^\infty$-direct sum of two Hilbert spaces $\mathbb{C}$, whose base is denoted by $s_0$, and $\ell^2(\mathbb{N})$, whose base is denoted by $\{s_k\}_{k=1}^\infty$. We define an inner product $\langle \cdot, \cdot \rangle_X : X \times X \to A$ by $\langle s_0, s_0 \rangle_X = p_0$, $\langle s_k, s_\ell \rangle_X = p_1$ for $k, \ell = 1, 2, \ldots$, and $\langle s_k, s_\ell \rangle_X = 0$ for $k \neq \ell$. The right action of $A$ on $X$ is defined by

$$s_k p_i = \begin{cases} s_0 & \text{for } k = i = 0, \\ s_k & \text{for } k \geq 1, i = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then $X$ becomes a Hilbert $A$-module. We define a left action $\varphi_X : A \to \mathcal{L}(X)$ by $\varphi_X(p_0) = \varphi_X(p_1) = 0$, and $\varphi_X(p_2) = \text{id}_X$. Now we get a $C^*$-correspondence $X$ over $A$. This $C^*$-correspondence is defined from the following graph;

(see [Katsura 2004a]). Let us define ideals of $A$ by

$$I_0 = \mathbb{C}p_0, \ I_1 = \mathbb{C}p_1, \ I_{01} = \mathbb{C}p_0 + \mathbb{C}p_1 \ \text{and} \ \ I_{12} = \mathbb{C}p_1 + \mathbb{C}p_2.$$
Since \( \ker \varphi_X = \varphi_X^{-1}(\mathcal{H}(X)) = I_{01} \), we have \( J_X = 0 \). Hence all ideals are negatively invariant. Since \( X(I_1) = X(I_{01}) = 0 \), both \( I_1 \) and \( I_{01} \) are invariant. By straightforward computation, we get \( J(I_1) = I_{12} \) and \( J(I_{01}) = I_{01} \). Thus two ideals \( I_1, I_{01} \) satisfy that \( I_1 \subset I_{01} \) and \( J(I_1) \not\subset J(I_{01}) \). We can see that \( \mathcal{O}_X \) is isomorphic to the direct sum of \( M_2(\mathbb{C}) \) and the unitization \( \tilde{K} \) of the \( C^* \)-algebra \( K \) of compact operators on \( \ell^2(\mathbb{N}) \). There exist six \( O \)-pairs (see Definition 5.12) which correspond to six ideals of \( \mathcal{O}_X \cong M_2(\mathbb{C}) \oplus \tilde{K} \):

\[
(0, 0) \subset (I_1, I_1) \subset (I_1, I_{12}) \quad 0 \subset K \subset \tilde{K}
\]

\[
(I_0, I_0) \subset (I_{01}, I_{01}) \subset (A, A) \quad M_2(\mathbb{C}) \subset M_2(\mathbb{C}) \oplus K \subset \mathcal{O}_X.
\]

This example also shows that \( J(I_1 \cap I_2) \subset J(I_1) \cap J(I_2) \) does not hold in general for two ideals \( I_1, I_2 \) of \( A \). However, the converse inclusion \( J(I_1) \cap J(I_2) \subset J(I_1 \cap I_2) \) always holds.

**Proposition 5.5.** For two ideals \( I_1, I_2 \) of \( A \), we have \( J(I_1) \cap J(I_2) \subset J(I_1 \cap I_2) \).

**Proof.** Take \( a \in J(I_1) \cap J(I_2) \). Since \( [\varphi_X(a)]_{I_1} \in \mathcal{H}(X_{I_1}) \) and \( [\varphi_X(a)]_{I_2} \in \mathcal{H}(X_{I_2}) \), we have \([\varphi_X(a)]_{I_1 \cap I_2} \in \mathcal{H}(X_{I_1 \cap I_2})\) by Corollary 1.12. We get \( aX^{-1}(I_1 \cap I_2) \subset I_1 \cap I_2 \) from

\[
aX^{-1}(I_1 \cap I_2) \subset aX^{-1}(I_1) \subset I_1, \quad aX^{-1}(I_1 \cap I_2) \subset aX^{-1}(I_2) \subset I_2.
\]

Hence \( a \in J(I_1 \cap I_2) \). Thus we have \( J(I_1) \cap J(I_2) \subset J(I_1 \cap I_2) \). \( \square \)

**Definition 5.6.** Let \( X \) be a \( C^* \)-correspondence over a \( C^* \)-algebra \( A \). A \textit{T-pair} of \( X \) is a pair \( \omega = (I, I') \) of ideals \( I, I' \) of \( A \) such that \( I \) is positively invariant and \( I \subset I' \subset J(I) \).

**Definition 5.7.** Let \( \omega_1 = (I_1, I_1') \) and \( \omega_2 = (I_2, I_2') \) be \( T \)-pairs. We write \( \omega_1 \subset \omega_2 \) if \( I_1 \subset I_2 \) and \( I_1' \subset I_2' \). We denote \( \omega_1 \cap \omega_2 \) the pair \((I_1 \cap I_2, I_1' \cap I_2')\).

**Proposition 5.8.** For two \( T \)-pairs \( \omega_1 = (I_1, I_1'), \omega_2 = (I_2, I_2'), \) their intersection \( \omega_1 \cap \omega_2 = (I_1 \cap I_2, I_1' \cap I_2') \) is a \( T \)-pair.

**Proof.** By Proposition 4.10, \( I_1 \cap I_2 \) is a positively invariant ideal. By Proposition 5.5, we have

\[
I_1 \cap I_2 \subset I_1' \cap I_2' \subset J(I_1) \cap J(I_2) \subset J(I_1 \cap I_2).
\]

Hence \( \omega_1 \cap \omega_2 \) is a \( T \)-pair. \( \square \)

\( T \)-pairs arise from representations.

**Definition 5.9.** For a representation \( (\pi, t) \) of \( X \), we define \( I_{(\pi, t)}, I'_{(\pi, t)} \subset A \) by

\[
I_{(\pi, t)} = \ker \pi, \quad I'_{(\pi, t)} = \pi^{-1}(\psi_t(\mathcal{H}(X))).
\]

The pair \((I_{(\pi, t)}, I'_{(\pi, t)})\) is denoted by \( \omega_{(\pi, t)} \).
Clearly \( I_{\pi,t} \) is an ideal of \( A \). By the remark before Definition 2.8, we see that \( I'_{\pi,t} \) is also an ideal of \( A \).

**Lemma 5.10.** For a representation \((\pi,t)\) of a \( C^*\)-correspondence \( X \) over a \( C^*\)-algebra \( A \), we have the following.

(i) \( I_{\pi,t} \) is positively invariant.

(ii) \( \ker t = X I_{\pi,t} \).

(iii) There exists an injective representation \((\tilde{\pi}, i)\) of the \( C^*\)-correspondence \( X I_{\pi,t} \) on \( C^* (\tilde{\pi}, t) \) such that \((\pi,t) = (\tilde{\pi} \circ [\cdot], i \circ [\cdot] I_{\pi,t}) \).

(iv) \( a \in I'_{\pi,t} \) implies \( [\varphi_X(a)] I_{\pi,t} \in \mathcal{H}(X I_{\pi,t}) \) and \( \pi(a) = \psi_I([\varphi_X(a)] I_{\pi,t}) \).

(v) For an element \( a \in A \) with \( \varphi_X(a) \in \mathcal{H}(X) \), we have \( \pi(a) = \psi_I(\varphi_X(a)) \) if and only if \( a \in I'_{\pi,t} \).

**Proof.**

(i) For \( a \in I_{\pi,t} \) and \( \xi, \eta \in X \), we have \( \langle \eta, \varphi_X(a) \xi \rangle \in I_{\pi,t} \) because

\[
\pi(\langle \eta, \varphi_X(a) \xi \rangle) = t(\eta)^* t(\varphi_X(a)) \xi = t(\eta)^* \pi(a) t(\xi) = 0.
\]

Hence \( X(I_{\pi,t}) \subseteq I_{\pi,t} \). Thus \( I_{\pi,t} \) is positively invariant.

(ii) For \( \xi \in X \), we have

\[
\xi \in \ker t \iff t(\xi) = 0 \iff t(\xi)^* t(\xi) = 0 \iff \pi(\langle \xi, \xi \rangle) = 0 \iff \langle \xi, \xi \rangle \in I_{\pi,t} \iff \xi \in X I_{\pi,t}.
\]

(iii) Obvious by the definition of \( I_{\pi,t} \) and (ii).

(iv) Since \( a \in I'_{\pi,t} \), we can find \( k \in \mathcal{H}(X) \) with \( \pi(a) = \psi_I(k) \). For \( \xi \in X \), we have

\[
t(\varphi_X(a) \xi) = \pi(a) t(\xi) = \psi_I(k) t(\xi) = t(k \xi).
\]

Hence \( (\varphi_X(a) - k) \xi \in \ker t = X I_{\pi,t} \) for all \( \xi \in X \). This implies that \( [\varphi_X(a)] I_{\pi,t} = [k] I_{\pi,t} \in \mathcal{H}(X I_{\pi,t}) \) and

\[
\pi(a) = \psi_I(k) = \psi_I([k] I_{\pi,t}) = \psi_I([\varphi_X(a)] I_{\pi,t}).
\]

(v) If \( \pi(a) = \psi_I(\varphi_X(a)) \), then \( a \in I'_{\pi,t} \). For \( a \in I'_{\pi,t} \) with \( \varphi_X(a) \in \mathcal{H}(X) \), we have \( \pi(a) = \psi_I([\varphi_X(a)] I_{\pi,t}) = \psi_I(\varphi_X(a)) \) by (iv).

\( \square \)

**Proposition 5.11.** For a representation \((\pi,t)\) of \( X \), the pair \( \omega_{\pi,t} \) is a \( T \)-pair.

**Proof.** By Lemma 5.10 (i), \( I_{\pi,t} \) is positively invariant. Clearly we have \( I_{\pi,t} \subseteq I'_{\pi,t} \). Take \( a \in I'_{\pi,t} \). We have \( [\varphi_X(a)] I_{\pi,t} \in \mathcal{H}(X I_{\pi,t}) \) by Lemma 5.10 (iv). Take \( b \in X^{-1}(I_{\pi,t}) \). Since \( ab \in I'_{\pi,t} \), we have \( \pi(ab) = \psi_I([\varphi_X(ab)] I_{\pi,t}) \) by Lemma 5.10 (iv). We see \( [\varphi_X(ab)] I_{\pi,t} = 0 \) because \( ab \in X^{-1}(I_{\pi,t}) \). Hence \( \pi(ab) = 0 \). Thus we get \( ab \in \ker \pi = I_{\pi,t} \). This shows \( a \in J(I_{\pi,t}) \). Hence we get \( I'_{\pi,t} \subseteq J(I_{\pi,t}) \). Thus \( \omega_{\pi,t} = (I_{\pi,t}, I'_{\pi,t}) \) is a \( T \)-pair.

\( \square \)
We will see that all $T$-pairs come from representations (Proposition 6.12). In the same way as in the proof of Proposition 5.11, we can see that for a morphism $(\Pi, T)$ from a $C^*$-correspondence $X$ to a $C^*$-correspondence $Y$, the pair $\omega_{(\Pi, T)} = (I_{(\Pi, T)}, I'_{(\Pi, T)})$ defined by

$$I_{(\Pi, T)} = \ker \Pi, \quad I'_{(\Pi, T)} = (\varphi_Y \circ \Pi)^{-1}\left(\Psi_T(\mathcal{O}(X))\right)$$

is a $T$-pair.

**Definition 5.12.** A $T$-pair $\omega = (I, I')$ satisfying $J_X \subset I'$ is called an $O$-pair.

It is clear that the intersection $\omega_1 \cap \omega_2$ of two $O$-pairs $\omega_1, \omega_2$ is an $O$-pair.

**Lemma 5.13.** A pair $\omega = (I, I')$ of ideals of $A$ is an $O$-pair if and only if $I$ is invariant and $I + J_X \subset I' \subset J(I)$.

**Proof.** For an $O$-pair $\omega = (I, I')$, we have $I + J_X \subset I' \subset J(I)$. Thus we get $J_X \subset J(I)$. Now Proposition 5.3 implies that $I$ is negatively invariant. Therefore $I$ is an invariant ideal. The converse is obvious. □

For a $C^*$-correspondence $X = C_\delta(E^1)$ arising from a topological graph $E$, an $O$-pair $(I, I')$ is in the form $(C_0(E^0 \setminus X^0), C_0(E^0 \setminus Z))$ where $(X^0, Z)$ is an admissible pair of closed sets of $E^0$ defined in [Katsura 2006a].

**Proposition 5.14.** A representation $(\pi, t)$ is covariant if and only if the pair $\omega_{(\pi, t)}$ is an $O$-pair.

**Proof.** If $(\pi, t)$ is covariant, then clearly $J_X \subset I'_{(\pi, t)}$. Thus $\omega_{(\pi, t)}$ is an $O$-pair. Conversely, if $\omega_{(\pi, t)}$ is an $O$-pair, then for $a \in J_X \subset I'_{(\pi, t)}$, we have $\pi(a) = \psi_t(\varphi_X(a))$ by Lemma 5.10 (v). Hence $(\pi, t)$ is covariant. □

By Proposition 5.14, we have $\omega_{(\pi, t)} = (0, J_X)$ for all injective covariant representations $(\pi, t)$.

### 6. $C^*$-correspondences associated with $T$-pairs

Take a $T$-pair $\omega = (I, I')$ of $X$ and fix it throughout this section. In this section, we construct a $C^*$-algebra $A_\omega$, a $C^*$-correspondence $X_\omega$ over $A_\omega$ and a representation $(\pi_\omega, t_\omega)$ of $X$ on the $C^*$-algebra $\mathcal{O}_{X_\omega}$. In the next section, we will see that this representation $(\pi_\omega, t_\omega)$ has a universal property.

![Diagram](attachment:image.png)
**Definition 6.1.** For a $T$-pair $\omega = (I, I')$ of a $C^*$-correspondence $X$ over $A$, we define a $C^*$-algebra $A_\omega$ and a Hilbert $A_\omega$-module $X_\omega$ by

$$A_\omega = \{(b, b') \in A/I \oplus A/I' \mid [b]_{J(I)/I} = [b']_{J(I)/I'} \in A/J(I)\},$$

$$X_\omega = \{((\eta, \eta')) \in X_I \oplus X_{I'} \mid [\eta]_{J(I)/I} = [\eta']_{J(I)/I'} \in X_{J(I)}\},$$

where the operations are defined as in Section 1.

Note that $A_\omega$ is a pull-back $C^*$-algebra of two surjections $[\cdot]_{J(I)/I} : A/I \to A/J(I)$ and $[\cdot]_{J(I)/I'} : A/I' \to A/J(I)$.

**Definition 6.2.** We define a $*$-homomorphism $\Psi_\omega : \mathcal{L}(X_I) \to \mathcal{L}(X_\omega)$ by

$$\Psi_\omega(S)(\eta, \eta') = (S\eta, [S]_{I'/I}\eta') \in X_\omega$$

for $S \in \mathcal{L}(X_I)$ and $(\eta, \eta') \in X_\omega$.

**Definition 6.3.** We define a left action $\varphi_{X_\omega} : A_\omega \to \mathcal{L}(X_\omega)$ by

$$\varphi_{X_\omega}(b, b') = \Psi_\omega(\varphi_X(b)),$$

for $(b, b') \in A_\omega$. Thus $X_\omega$ is a $C^*$-correspondence over $A_\omega$.

**Definition 6.4.** We set

$$\Pi_\omega : A/I \ni b \mapsto (b, [b]_{I'/I}) \in A_\omega, \quad T_\omega : X_I \ni \eta \mapsto ([\eta]_{I'/I}) \in X_\omega.$$

**Lemma 6.5.** We have $\varphi_{X_\omega} \circ \Pi_\omega = \Psi_\omega \circ \varphi_X$, and $T_\omega(S\eta) = \Psi_\omega(S)T_\omega(\eta)$ for $S \in \mathcal{L}(X_I)$ and $\eta \in X_I$.

**Proof:** Clear by the definitions. □

From this lemma, we easily get the following.

**Proposition 6.6.** The pair $(\Pi_\omega, T_\omega)$ is an injective morphism from $X_I$ to $X_\omega$, and the map $\Psi_{T_\omega} : \mathcal{K}(X_I) \to \mathcal{K}(X_\omega)$ coincides with the restriction of $\Psi_\omega$ to $\mathcal{K}(X_I)$.

The next proposition is also easy to see from the definitions.

**Proposition 6.7.** For a $T$-pair $\omega = (I, I')$ with $I' = J(I)$, the morphism $(\Pi_\omega, T_\omega)$ from $X_I$ to $X_\omega$ is an isomorphism.

To compute $J_{X_\omega} \subseteq A_\omega$, we need the following lemma.

**Lemma 6.8.** A pair $(\Pi, T)$ of maps defined by

$$\Pi : A_\omega \ni (b, b') \mapsto b \in A/I, \quad T : X_\omega \ni (\eta, \eta') \mapsto \eta \in X_I,$$

is a morphism from $X_\omega$ to $X_I$ satisfying $\Pi \circ T_\omega = \text{id}_{A/I}$ and $T \circ T_\omega = \text{id}_{X_I}$. A $*$-homomorphism $\Psi : \mathcal{L}(X_\omega) \ni S \mapsto T \circ S \circ T_\omega \in \mathcal{L}(X_I)$ satisfies that $\Psi \circ \Psi_\omega = \text{id}_{\mathcal{L}(X_I)}$ and the restriction of $\Psi$ to $\mathcal{K}(X_\omega)$ coincides with $\Psi_T : \mathcal{K}(X_\omega) \to \mathcal{K}(X_I)$. 

Proof. It is clear that \((\mathcal{T}, T)\) is a morphism satisfying \(\Pi \circ \Pi_\omega = \text{id}_{A/I}\) and \(T \circ T_\omega = \text{id}_{X_I}\). By Lemma 6.5, we have

\[
\Psi(\Psi_\omega(S)) = T(\Psi_\omega(S)T_\omega(\eta)) = T(T_\omega(S\eta)) = S\eta,
\]

for \(S \in \mathcal{L}(X_I)\) and \(\eta \in X_I\). This proves \(\Psi \circ \Psi_\omega = \text{id}_{\mathcal{L}(X_I)}\). For \((\eta_1, \eta'_1), (\eta_2, \eta'_2) \in X_\omega\) and \(\eta \in X_I\), we have

\[
\Psi(\theta_{(\eta_1, \eta'_1), (\eta_2, \eta'_2)} T_\omega(\eta)) = T((\eta_1(\eta_2, \eta)_X, \eta'_1(\eta'_2, [\eta]_{I'/I})_{X_I})) = \eta_1(\eta_2, \eta)_X = \theta_{\eta_1, \eta_2}(\eta).
\]

Hence we have \(\Psi(\theta_{(\eta_1, \eta'_1), (\eta_2, \eta'_2)} ) = \theta_{\eta_1, \eta_2}\). This shows that the restriction of \(\Psi\) to \(\mathcal{H}(X_\omega)\) coincides with \(\mathcal{H}(X_I)\).

\[\blacksquare\]

Proposition 6.9. We have

\[
\ker \varphi_{X_\omega} = \{(b, b') \in A_\omega \mid b \in \ker \varphi_{X_I}\},
\]

\[
\varphi_{X_\omega}^{-1}(\mathcal{H}(X_\omega)) = \{(b, b') \in A_\omega \mid b \in \varphi_{X_I}^{-1}(\mathcal{H}(X_I))\},
\]

\[
J_{X_\omega} = \{(b, b') \in A_\omega \mid b \in J_{X_I}, b' = 0\}.
\]

Proof. Since \(\Psi \circ \Psi_\omega = \text{id}_{\mathcal{L}(X_I)}\) by Lemma 6.8, we have \(\Psi(\varphi_{X_\omega}((b, b'))) = \varphi_{X_I}(b)\) for \((b, b') \in A_\omega\). Hence for \((b, b') \in A_\omega\), we have that \(\varphi_{X_\omega}((b, b')) = 0\) if and only if \(\varphi_{X_I}(b) = 0\). This proves the first equality. The second one follows similarly because we have \(\Psi_\omega(\mathcal{H}(X_I)) \subset \mathcal{H}(X_\omega)\) and \(\Psi(\mathcal{H}(X_\omega)) \subset \mathcal{H}(X_I)\). We will prove the third equality. It is easy to see that for \(b \in J_{X_I}\), we have

\[
(b, 0) \in \varphi_{X_\omega}^{-1}(\mathcal{H}(X_\omega)) \cap (\ker \varphi_{X_\omega})^\perp = J_{X_\omega}.
\]

Take \((b, b') \in J_{X_\omega}\), and we will prove that \(b \in J_{X_I}\) and \(b' = 0\). Since \(\varphi_{X_\omega}((b, b')) \in \mathcal{H}(X_\omega)\), we have \(\varphi_{X_I}(b) \in \mathcal{H}(X_I)\). For any \(b_0 \in \ker \varphi_{X_I} \subset A/I\), we have \(\Pi_\omega(b_0) = (b_0, [b_0]_{I'/I}) \in \ker \varphi_{X_\omega}\). Hence \((b, b')(b_0, [b_0]_{I'/I}) = 0\). This implies that \(b \in (\ker \varphi_{X_I})^\perp\). Hence \(b \in J_{X_I}\). Since \(J_{X_I} = J(I)/I\) by Lemma 5.2, we have

\[
[b]_{J(I)/I'} = [b]_{J(I)/I} = 0.
\]

Hence \((0, b^{**}) \in A_\omega\). Since \((0, b^{**}) \in \ker \varphi_{X_\omega}\), we have \((b, b')(0, b^{**}) = 0\). This implies \(b' = 0\). Thus we get \(J_{X_\omega} = \{(b, b') \in A_\omega \mid b \in J_{X_I}, b' = 0\}\).

We have \(J_{X_\omega} = \{(b, b') \in A_\omega \mid b' = 0\}\) because for \(b \in A/I\) we have \((b, 0) \in A_\omega\) if and only if \(b \in J_{X_I}\).
Definition 6.10. We define a $\ast$-homomorphism $\pi_\omega : A \to \mathcal{O}_{X_\omega}$ and a linear map $t_\omega : X \to \mathcal{O}_{X_\omega}$ by
\[
\pi_\omega(a) = \pi_{X_\omega}(\Pi_\omega([a]_I)), \quad t_\omega(\xi) = t_{X_\omega}(T_\omega([\xi]_I))
\]
for $a \in A$ and $\xi \in X$, where $(\pi_{X_\omega}, t_{X_\omega})$ is the universal covariant representation of the $C^\ast$-correspondence $X_\omega$ on $\mathcal{O}_{X_\omega}$.

Proposition 6.11. The pair $(\pi_\omega, t_\omega)$ is a representation of $X$ on $\mathcal{O}_{X_\omega}$, which admits a gauge action and satisfies $C^\ast(\pi_\omega, t_\omega) = \mathcal{O}_{X_\omega}$.

Proof. Since $(\pi_\omega, t_\omega)$ is a composition of morphisms, it is a representation. Clearly the gauge action of $\mathcal{O}_{X_\omega}$ gives a gauge action for the representation $(\pi_\omega, t_\omega)$. We will prove $C^\ast(\pi_\omega, t_\omega) = \mathcal{O}_{X_\omega}$. Since $\mathcal{O}_{X_\omega}$ is generated by the images of $\pi_{X_\omega}$ and $t_{X_\omega}$, it suffices to show that
\[
\pi_{X_\omega}(A_\omega), t_{X_\omega}(X_\omega) \subseteq C^\ast(\pi_\omega, t_\omega).
\]
Take $(b, b') \in A_\omega$. Choose $a \in A$ with $[a]_I' = b'$. We have $b - [a]_I \in J(I)/I = J_{X_I}$ because $[b]_{J(I)/I} = [b']_{J(I)/I'} = [a]_{J(I)}$. Thus we have $\varphi_{X_I}(b - [a]_I) \in \mathcal{C}(X_I)$. Hence there exists $k \in \mathcal{C}(X)$ such that $[k]_I = \varphi_{X_I}(b - [a]_I)$. Since $(b - [a]_I, 0) \in J_{X_\omega}$ by Proposition 6.9, we have
\[
\pi_{X_\omega}((b - [a]_I, 0)) = \psi_{t_{X_\omega}}(\varphi_{X_\omega}((b - [a]_I, 0))) = \psi_{t_{X_\omega}}(\varphi_{X_\omega}(\psi_{X_I}(b - [a]_I)))
\]
\[
= \psi_{t_{X_\omega}}(\psi_{t_{\omega}}([k]_I)) = \psi_{t_\omega}(k).
\]
Therefore we get
\[
\pi_{X_\omega}((b, b')) = \pi_{X_\omega}(([a]_I, [a]_{I'})) + \pi_{X_\omega}((b - [a]_I, 0))
\]
\[
= \pi_\omega(a) + \psi_\omega(k) \in C^\ast(\pi_\omega, t_\omega).
\]
Thus we have shown that $\pi_{X_\omega}(A_\omega) \subseteq C^\ast(\pi_\omega, t_\omega)$.

Take $(\eta, \eta') \in X_\omega$. Choose $\xi \in X$ with $[\xi]_I = \eta'$. As above, we get $\eta - [\xi]_I \in X_I J_{X_I}$. Choose $\xi' \in X$ and $b \in J_{X_I}$ with $\eta - [\xi]_I = [\xi']_I b$. Then we have $(\eta - [\xi]_I, 0) = T_\omega([\xi']_I)(b, 0)$. Hence we get
\[
t_{X_\omega}((\eta, \eta')) = t_{X_\omega}(([\xi]_I, [\xi']_I)) + t_{X_\omega}((\eta - [\xi]_I, 0))
\]
\[
= t_\omega(\xi) + t_{X_\omega}(T_\omega([\xi']_I))\pi_{X_\omega}((b, 0))
\]
\[
= t_\omega(\xi) + t_\omega(\xi')\pi_{X_\omega}((b, 0)) \in C^\ast(\pi_\omega, t_\omega),
\]
because $\pi_{X_\omega}((b, 0)) \in C^\ast(\pi_\omega, t_\omega)$ as shown above. This completes the proof. 

Proposition 6.12. For a $T$-pair $\omega = (I, I')$, we have $\omega(\pi_\omega, t_\omega) = \omega$. 

Proof. Since the maps $\Pi_\omega: A_\tau \rightarrow A_\tau$ and $\pi_{X_\omega}: A_\tau \rightarrow \mathcal{C}_{X_\omega}$ are injective, we have

$$I_{(\pi_\omega, \pi_\omega)} = \ker \pi_\omega = \ker([\cdot]_I) = I.$$ 

For $a \in I'$, we have $[a]_I \in I'/I \subset J(I)/I = J_{X_\tau}$. Since $\Pi_\omega([a]_I) = ([a]_I, 0) \in J_{X_\tau}$, we have

$$\pi_\omega(a) = \pi_{X_\omega}(\Pi_\omega([a]_I)) = \psi_{1_{X_\omega}}(\varphi_{X_\omega}(([a]_I, 0))).$$

We see $\varphi_{X_\omega}([a]_I) \in \mathcal{H}(X_I)$ from $[a]_I \in J_{X_\tau}$. Hence by the definition of $\varphi_{X_\omega}$ we get

$$\varphi_{X_\omega}(([a]_I, 0)) = \Psi_{T_\omega}(\varphi_{X_\omega}([a]_I)) \in \Psi_{T_\omega}(\mathcal{H}(X_I)).$$

Since $\mathcal{H}(X_I) = [\mathcal{H}(X)_I]$, we have

$$\pi_\omega(a) \in \psi_{1_{X_\omega}}(\Psi_{T_\omega}([\mathcal{H}(X)_I])) = \psi_{T_\omega}(\mathcal{H}(X)).$$

Hence $a \in I'_{(\pi_\omega, \pi_\omega)}$. We have shown that $I' \subset I'_{(\pi_\omega, \pi_\omega)}$. Conversely take $a \in I'_{(\pi_\omega, \pi_\omega)}$. Since

$$\pi_{X_\omega}(\Pi_\omega([a]_I)) = \pi_\omega(a) \in \psi_{T_\omega}(\mathcal{H}(X)) \subset \psi_{1_{X_\omega}}(\mathcal{H}(X_\omega)),$$

we have $\Pi_\omega([a]_I) \in J_{X_\omega}$. Hence by Proposition 6.9, we have $[a]_{I'} = 0$. This means $a \in I'$. Thus we get $I'_{(\pi_\omega, \pi_\omega)} \subset I'$. Therefore $I'_{(\pi_\omega, \pi_\omega)} = I'$. We have shown that $\omega_{(\pi_\omega, \pi_\omega)} = \omega$.

By Proposition 6.12, we see that all T-pairs come from representations.

7. $C^*$-algebras generated by representations of $C^*$-correspondences

In this section, we prove the following theorem.

Theorem 7.1. Let $X$ be a $C^*$-correspondence over a $C^*$-algebra $A$, and $(\pi, t)$ be a representation of $X$. If a T-pair $\omega$ of $X$ satisfies $\omega \subset \omega_{(\pi, t)}$, then there exists a unique surjective $\ast$-homomorphism $\rho: \mathcal{C}_{X_\omega} \rightarrow C^*(\pi, t)$ such that $\pi = \rho \circ \pi_\omega$ and $t = \rho \circ t_\omega$. The surjection $\rho$ is an isomorphism if and only if $\omega = \omega_{(\pi, t)}$ and $(\pi, t)$ admits a gauge action.

Take a representation $(\pi, t)$ of a $C^*$-correspondence $X$ and a T-pair $\omega = (I, I')$ of $X$ satisfying $\omega \subset \omega_{(\pi, t)}$. In order to get a $\ast$-homomorphism $\rho: \mathcal{C}_{X_\omega} \rightarrow C^*(\pi, t)$, we will construct a covariant representation $(\tilde{\pi}, \tilde{t})$ of the $C^*$-correspondence $X_\omega$ on $C^*(\pi, t)$. Since $I \subset I_{(\pi, t)} = \ker \pi$, we can define a representation $(\tilde{\pi}, \tilde{t})$ of a $C^*$-correspondence $X_I$ over $A/I$ on $C^*(\pi, t)$ such that $\tilde{\pi}([a]_I) = \pi(a)$ for $a \in A$ and $\tilde{t}([\xi]_I) = t(\xi)$ for $\xi \in X$ as in Lemma 5.10 (iii). It is easy to see that $I_{(\tilde{\pi}, \tilde{t})} = I_{(\pi, t)}/I$ and $I'_{(\tilde{\pi}, \tilde{t})} = I'_{(\pi, t)}/I$. 

Definition 7.2. Let \( (b, b') \in A_\omega \). Take \( d \in A/I \) with \( [d]_{I/I} = b' \). Define \( \tilde{\pi}((b, b')) \in C^*(\pi, t) \) by
\[
\tilde{\pi}((b, b')) = \hat{\pi}(d) + \psi_t(\varphi_{X_I}(b - d)) \in C^*(\pi, t).
\]

Note that this definition makes sense because \( b - d \in J(I)/I = J_{X_I} \) implies \( \varphi_{X_I}(b - d) \in \mathcal{H}(X_I) \). Note also that \( \tilde{\pi}((b, b')) \in C^*(\pi, t) \) does not depend on the choice of \( d \in A/I \) with \( [d]_{I/I} = b' \) because we have \( \hat{\pi}(d_1 - d_2) = \psi_t(\varphi_{X_I}(d_1 - d_2)) \) if \( d_1 - d_2 \in I'/I \subset I_{(\pi, t)}/I = I'_{(\tilde{\pi}, \hat{\pi})} \) by Lemma 5.10 (v).

Lemma 7.3. The map \( \tilde{\pi} : A_\omega \to C^*(\pi, t) \) is a \(*\)-homomorphism.

Proof. It is obvious that \( \tilde{\pi} \) is a \(*\)-preserving linear map. We will show \( \tilde{\pi} \) is multiplicative. Take \((b_1, b'_1), (b_2, b'_2) \in A_\omega \). Take \( d_1, d_2 \in A/I \) with \( [d_1]_{I/I} = b'_1, [d_2]_{I/I} = b'_2 \). Since
\[
\hat{\pi}(d)\psi_t(\varphi_{X_I}(b)) = \psi_t(\varphi_{X_I}(db))
\]
for \( d \in A/I \) and \( b \in J(I)/I = J_{X_I} \), we have
\[
\tilde{\pi}((b_1, b'_1))\tilde{\pi}((b_2, b'_2))
\]
\[
= (\hat{\pi}(d_1) + \psi_t(\varphi_{X_I}(b_1 - d_1))) (\hat{\pi}(d_2) + \psi_t(\varphi_{X_I}(b_2 - d_2)))
\]
\[
= \hat{\pi}(d_1d_2) + \psi_t(\varphi_{X_I}(d_1(b_2 - d_2) + (b_1 - d_1)d_2 + (b_1 - d_1)(b_2 - d_2)))
\]
\[
= \hat{\pi}(d_1d_2) + \psi_t(\varphi_{X_I}(b_1b_2 - d_1d_2))
\]
\[
= \tilde{\pi}((b_1b_2, b'_1b'_2))
\]
\[
= \tilde{\pi}((b_1, b'_1)(b_2, b'_2)).
\]
Hence \( \tilde{\pi} \) is a \(*\)-homomorphism. \( \square \)

Proposition 7.4. The map \( \tilde{\pi} : A_\omega \to C^*(\pi, t) \) is injective if and only if \( \omega = \omega_{(\pi, t)} \).

Proof. Suppose that \( \tilde{\pi} \) is injective. For \( a \in I_{(\pi, t)} \), we have \( ([a]_I, [a]_{I'}) \in A_\omega \) and
\[
\tilde{\pi}(([a]_I, [a]_{I'})) = \hat{\pi}([a]_I) = \pi(a) = 0.
\]
Hence \( ([a]_I, [a]_{I'}) = 0 \). This implies \( a \in I \). Thus we get \( I_{(\pi, t)} = I \). For \( a \in I'_{(\pi, t)} \), we have \([a]_I \in I'_{(\pi, t)}/I \subset J(I_{(\pi, t)})/I = J(I)/I \). Hence we get \( 0, [a]_{I'} \in A_\omega \). We
also get \( \varphi_{X_i}([a]_t) \in \mathcal{K}(X_I) \). Since \([a]_I \in I'_{(\pi,t)}' \) and \( I = I'_{(\tilde{\pi},i)}' \), we have

\[
\tilde{\pi}((0, [a]_I)) = \tilde{\pi}([a]_I) - \psi_i(\varphi_{X_i}([a]_I)) = 0,
\]

by Lemma 5.10 (v). Since \( \tilde{\pi} \) is injective, we have \((0, [a]_I) = 0\). This implies \( a \in I' \). Thus we get \( I'_{(\pi,t)} = I' \). Therefore if \( \tilde{\pi} \) is injective, then \( \omega = \omega_{(\pi,t)} \).

Conversely assume \( \omega = \omega_{(\pi,t)} \). Take \((b, b') \in A_\omega\) with \( \tilde{\pi}((b, b')) = 0\). Take \( d \in A/I\) with \([d]_{I'_{(\pi,t)}} = b'\). Then we have \( \tilde{\pi}(d) = \psi_i(\varphi_{X_i}(d - b)) \). Hence \( d \in I'_{(\pi,t)} \). Therefore we have \( b' = 0 \). We also have \( \psi_i(\varphi_{X_i}(b)) = 0 \). Since \( I = I'_{(\pi,t)} \), the map \( i \) is injective. Hence \( \psi_i \) is also injective. Therefore we have \( b \in \ker \varphi_{X_i} \). We also have \( b \in J(I)/I = J_{X_I} \) because \([b]_{J(I)/I} = [b']_{J(I)/I'} = 0 \). Hence \( b = 0 \). We have proved that \( \tilde{\pi} \) is injective.

**Definition 7.5.** Let \( \zeta \in X_I J_{X_i} \). Take \( \eta \in X_I \) and \( b \in J_{X_i} \) such that \( \zeta = \eta b \). We define \( \bar{i}(\xi) = i(\eta)\psi_i(\varphi_{X_i}(b)) \) for all \( \xi \in C^*(\pi, t) \).

**Lemma 7.6.** The map \( i: X_I J_{X_i} \rightarrow C^*(\pi, t) \) is a well-defined linear map satisfying that \( i(\eta)^*\bar{i}(\xi) = \psi_i(\varphi_{X_i}(\langle \eta, \zeta \rangle_{X_{ij}})) \) for all \( \zeta \in X_I J_{X_i} \) and \( \eta \in X_I \) and \( i(\xi_1)^*\bar{i}(\xi_2) = \psi_i(\varphi_{X_i}(\langle \xi_1, \xi_2 \rangle_{X_{ij}})) \) for all \( \xi_1, \xi_2 \in X_I J_{X_i} \).

**Proof.** Take \( \eta_1, \eta_2 \in X_I \) and \( b_1, b_2 \in J_{X_i} \), and define \( \xi_1, \xi_2 \in X_I J_{X_i} \) by \( \xi_1 = \eta_1 b_1 \) and \( \xi_2 = \eta_2 b_2 \). We have

\[
i(\eta_1)^*i(\eta_2)\psi_i(\varphi_{X_i}(b_2)) = \tilde{\pi}((\eta_1, \eta_2)_{X_{ij}})\psi_i(\varphi_{X_i}(b_2)) = \psi_i(\varphi_{X_i}(\langle \eta_1, \eta_2 \rangle_{X_{ij}})) = \psi_i(\varphi_{X_i}(\langle \eta_1, \xi_2 \rangle_{X_{ij}})).
\]

A similar computation shows that

\[
i(\eta_1)^*i(\eta_2)\psi_i(\varphi_{X_i}(b_1)) = \psi_i(\varphi_{X_i}(\langle \xi_1, \xi_2 \rangle_{X_{ij}}))
\]

For \( \zeta \in X_I J_{X_i} \), take \( \eta_1, \eta_2 \in X_I \) and \( b_1, b_2 \in J_{X_i} \) such that \( \zeta = \eta_1 b_1 = \eta_2 b_2 \). Set \( x = i(\eta_1)\psi_i(\varphi_{X_i}(b_1)) - i(\eta_2)\psi_i(\varphi_{X_i}(b_2)) \in C^*(\pi, t) \). We have \( x^*x = 0 \) because for \( i, j = 1, 2 \),

\[
i(\xi_j)\psi_i(\varphi_{X_i}(b_1)) = \psi_i(\varphi_{X_i}(\langle \xi_j, \xi \rangle_{X_{ij}})).
\]

This shows \( \bar{i} \) is well-defined. We can check the linearity of \( \bar{i} \) in a similar fashion. The two equalities in the statement had been already checked.

**Lemma 7.7.** We have

\[
\tilde{\pi}(b)\bar{i}(\xi) = \tilde{i}(\varphi_{X_i}(b)\xi), \quad \psi_i(k)\bar{i}(\xi) = \tilde{i}(k\xi),
\]

for \( b \in A/I, k \in \mathcal{K}(X_I) \), and \( \xi \in X_I J_{X_i} \).
Proof. Take \( \eta \in X_I \) and \( d \in J_{X_I} \) with \( \zeta = \eta d \). Then we have
\[
\tilde{\pi}(b)\tilde{\iota}(\zeta) = \tilde{\pi}(b)\tilde{\iota}(\eta)\psi_i(\varphi_{X_I}(d)) = \psi_i(k)\tilde{\iota}(\eta)\psi_i(\varphi_{X_I}(d)) = i(\varphi_{X_I}(b)\eta)d = i(\varphi_{X_I}(b)\eta)d = \tilde{\iota}(\varphi_{X_I}(b)\eta),
\]
we have
\[
\tilde{\iota}(\zeta) = i(\eta)\psi_i(\varphi_{X_I}(b)) = i(\eta)\tilde{\iota}(b) = i(\eta b) = i(\zeta).
\]
\[\square\]

Lemma 7.8. For \( \zeta \in X_I(I'/I) \), we have \( \tilde{\iota}(\zeta) = i(\zeta) \).

Proof. Choose \( \eta \in X_I \) and \( b \in I'/I \subset J(I)/I = J_{X_I} \) such that \( \zeta = \eta b \). Since \( b \in I'/I \subset I'(\pi,\tilde{\iota})/I = I'(\pi,\tilde{\iota}) \), we have \( \tilde{\pi}(b) = \psi_i(\varphi_{X_I}(b)) \) by Lemma 5.10 (v). Hence, we get
\[
\tilde{\iota}(\zeta) = i(\eta)\psi_i(\varphi_{X_I}(b)) = i(\eta)\tilde{\iota}(b) = i(\eta b) = i(\zeta).
\]
\[\square\]

Definition 7.9. Let \((\eta, \eta') \in X_{\omega} \). Take \( \zeta \in X_I \) such that \([\zeta]_{I'/I} = \eta' \). Define \( \tilde{\iota}((\eta, \eta')) \in C^*(\pi, t) \) by
\[
\tilde{\iota}((\eta, \eta')) = i(\zeta) + \tilde{\iota}(\eta - \zeta) \in C^*(\pi, t).
\]

Note that \( \eta - \zeta \in X_I J_{X_I} \), and that \( \tilde{\iota} : X_{\omega} \to C^*(\pi, t) \) is a well-defined linear map by Lemma 7.8.

Proposition 7.10. The pair \((\tilde{\pi}, \tilde{\iota})\) is a representation of the \( C^* \)-correspondence \( X_{\omega} \) on \( C^*(\pi, t) \) such that \( \tilde{\pi} = \tilde{\pi} \circ \Pi_{\omega} \) and \( \tilde{\iota} = \tilde{\iota} \circ T_{\omega} \).

Proof. It is easy to see that \( \tilde{\pi} = \tilde{\pi} \circ \Pi_{\omega} \) and \( \tilde{\iota} = \tilde{\iota} \circ T_{\omega} \). We will check that the pair \((\tilde{\pi}, \tilde{\iota})\) satisfies the two conditions in Definition 2.7. Take \((\eta_1, \eta'_1), (\eta_2, \eta'_2) \in X_{\omega} \). Choose \( \xi_1, \xi_2 \in X_I \) with \([\xi_1]_{I'/I} = \eta'_1, [\xi_2]_{I'/I} = \eta'_2 \). By Lemma 7.6, we have
\[
\tilde{\iota}((\eta_1, \eta'_1))^*\tilde{\iota}((\eta_2, \eta'_2)) = (\tilde{\iota}(\xi_1) + \tilde{\iota}(\eta_1 - \xi_1))^* (\tilde{\iota}(\xi_2) + \tilde{\iota}(\eta_2 - \xi_2))
\]
\[
= \tilde{\pi}(\langle \xi_1, \xi_2 \rangle_{X_I}) + \psi_i(\varphi_X(\langle \xi_1, \eta_2 - \xi_2 \rangle_{X_I} + \langle \eta_1 - \xi_1, \xi_2 \rangle_{X_I} + \langle \eta_1 - \xi_1, \eta_2 - \xi_2 \rangle_{X_I}))
\]
\[
= \tilde{\pi}(\langle \xi_1, \xi_2 \rangle_{X_I}) + \psi_i(\varphi_X(\langle \eta_1, \eta_2 \rangle_{X_I} - \langle \xi_1, \xi_2 \rangle_{X_I}))
\]
\[
= \tilde{\pi}(\langle \eta_1, \eta'_1 \rangle_{X_I}, \langle \eta_2, \eta'_2 \rangle_{X_I}).
\]

This proves condition (i) in Definition 2.7. We check condition (ii). Take \((b, b') \in A_{\omega} \) and \((\eta, \eta') \in X_{\omega} \). Choose \( d \in A/I \) and \( \zeta \in X_I \) with \([d]_{I'/I} = b' \) and \([\zeta]_{I'/I} = \eta' \).
By Lemma 7.7, we have
\[
\tilde{\pi}((b, b')) \tilde{i}((\eta, \eta')) = (\pi (d) + \psi_i(\varphi_{X_1}(b - d))) (i(\xi) + i(\eta - \zeta))
\]
\[
= \tilde{\pi}(d) i(\xi) + \psi_i(\varphi_{X_1}(b - d)) i(\xi)
+ \tilde{\pi}(d) i(\eta - \zeta) + \psi_i(\varphi_{X_1}(b - d)) i(\eta - \zeta)
\]
\[
= i(\varphi_{X_1}(d)) i + i(\varphi_{X_1}(b - d)) i(\eta - \zeta) + i(\varphi_{X_1}(d)) (\eta - \zeta)
\]
\[
= i(\varphi_{X_1}(b)) i(\eta - \zeta)).
\]

On the other hand, we have
\[
\varphi_{X_\omega}(b, b')(\eta, \eta') = (\varphi_{X_1}(b) \eta, [\varphi_{X_1}(b)]_I \eta') = (\varphi_{X_1}(b) \eta, [\varphi_{X_1}(b) \xi]_I / I).
\]

Hence we get
\[
\tilde{i}(\varphi_{X_\omega}(b, b')(\eta, \eta')) = i(\varphi_{X_1}(b) \eta) + i(\varphi_{X_1}(b) (\eta - \zeta)).
\]

Thus we have \(\tilde{\pi}((b, b')) \tilde{i}((\eta, \eta')) = \tilde{i}(\varphi_{X_\omega}(b, b')(\eta, \eta')).\) We are done. \(\square\)

**Proposition 7.11.** The representation \((\tilde{\pi}, \tilde{i})\) is covariant.

**Proof.** Take \((b, 0) \in J_{X_\omega}.\) By definition, we have \(\tilde{\pi}((b, 0)) = \psi_i(\varphi_{X_1}(b)).\) Since \(\varphi_{X_\omega}(b, 0) = \Psi_{T_0}(\varphi_{X_1}(b)),\) we have
\[
\psi_i(\varphi_{X_\omega}(b, 0)) = \psi_i(\psi_{\omega T_0}(\varphi_{X_1}(b))) = \psi_{\psi_{\omega T_0}}(\varphi_{X_1}(b)) = \psi_i(\varphi_{X_1}(b)).
\]

Hence we get \(\tilde{\pi}((b, 0)) = \psi_i(\varphi_{X_\omega}(b, 0))\) for every element \((b, 0) \in J_{X_\omega}.\) This completes the proof. \(\square\)

**Lemma 7.12.** The representation \((\tilde{\pi}, \tilde{i})\) of \(X_\omega\) is injective if and only if \(\omega = \omega_{(\pi, t)}//.\) It admits a gauge action if and only if \(\phi(\pi, t).\)

**Proof.** The first assertion follows from Proposition 7.4. If a representation \((\pi, t)\) admits a gauge action \(\beta,\) then \(\beta\) is also a gauge action for the representation \((\tilde{\pi}, \tilde{i})\) because \(\beta_i(\psi_i(k)) = \psi_i(k)\) for all \(k \in \mathfrak{A}(X)\) and \(z \in I.\) The converse is obvious. \(\square\)

Now we are ready to prove the main theorem of this section.

**Proof of Theorem 7.1.** Define \(\rho = \rho_{(\tilde{T}, \tilde{i})}: C_{X_\omega} \rightarrow C^*(\pi, t).\) Since \(\tilde{T} = \tilde{\pi} \circ \Pi_{\omega} \circ T_{\omega},\) we have \(\pi = \rho \circ \pi_{\omega}\) and \(t = \rho \circ t_{\omega}.\) This implies that \(\rho\) is surjective. The uniqueness follows from \(C^*(\pi_{\omega}, t_{\omega}) = C_{X_\omega}\) which was proved in Proposition 6.11. Finally by Lemma 7.12 and Theorem 3.6, \(\rho\) is an isomorphism if and only if \(\omega = \omega_{(\pi, t)}//.\) and \((\pi, t)\) admits a gauge action. \(\square\)

**Corollary 7.13.** Let \(X\) be a \(C^*-\)correspondence over a \(C^*-\)algebra \(A\) and \((\pi, t)\) be a representation of \(X\) which admits a gauge action. Then the \(C^*-\)algebra \(C^*(\pi, t)\) is naturally isomorphic to the \(C^*-\)algebra \(C_{X_{\omega_{(\pi, t)}}}^{\omega_{(\pi, t)}}.\)
We finish this section with a characterization of the $C^*$-algebra $\mathcal{C}_X$ without using $J_X$ and the notion of covariance.

**Proposition 7.14.** If a representation $(\pi, t)$ is injective and admits a gauge action, then there exists a surjection $\overline{\rho} : C^*(\pi, t) \to \mathcal{C}_X$ with $\pi X = \overline{\rho} \circ \pi$ and $t_X = \overline{\rho} \circ t$.

**Proof.** Set $\omega = \omega_{(\pi, t)} = (I_{(\pi, t)}, I'_{(\pi, t)})$. Since $(\pi, t)$ is injective, we have $I_{(\pi, t)} = 0$ and $I'_{(\pi, t)} \subset J(0) = J_X$. Hence we get $\omega \subset (0, J_X) = \omega_{(\pi, t, X)}$. Thus by Theorem 7.1, there exists a surjective *-homomorphism $\rho : \mathcal{C}_{X_{\omega}} \to \mathcal{C}_X$ with $\pi X = \rho \circ \pi_{\omega}$ and $t_X = \rho \circ t_{\omega}$. Since $(\pi, t)$ admits a gauge action, the $C^*$-algebra $C^*(\pi, t)$ is isomorphic to $\mathcal{C}_{X_{\omega}}$ by Corollary 7.13. This completes the proof. □

By Proposition 7.14, we can define $\mathcal{C}_X$ to be the smallest $C^*$-algebra among $C^*$-algebras generated by injective representations admitting gauge actions. Theorem 3.6 tells us that the covariance of representations characterizes the representation $(\pi_X, t_X)$ among injective representations admitting gauge actions.

**8. Structure of gauge-invariant ideals of $\mathcal{C}_X$**

We say that an ideal of $\mathcal{C}_X$ is gauge-invariant if it is globally invariant under the gauge action $\gamma$. In this section, we analyze structure of gauge-invariant ideals of $\mathcal{C}_X$.

**Definition 8.1.** For an ideal $P$ of $\mathcal{C}_X$, we define $I_P, I'_P \subset A$ by

$$\pi_X(I_P) = \pi_X(A) \cap P, \quad \pi_X(I'_P) = \pi_X(A) \cap (P + \psi_X(\mathfrak{A}(X))).$$

We set $\omega_P = (I_P, I'_P)$.

**Proposition 8.2.** For an ideal $P$ of $\mathcal{C}_X$, denote by $\sigma_P$ a natural surjection from $\mathcal{C}_X$ to $\mathcal{C}_X / P$. Then we have $\omega_P = \omega_{(\sigma_P \circ \pi_X, \sigma_P \circ t_X)}$. Hence $\omega_P$ is an O-pair.

**Proof.** Clear by the definitions. □

**Definition 8.3.** Let $\omega$ be an O-pair of $X$. The representation $(\pi_{\omega}, t_{\omega})$ of $X$ on $\mathcal{C}_{X_{\omega}}$ is covariant by Proposition 5.14 and Proposition 6.12. Hence there exists a surjection $\rho_{(\pi_{\omega}, t_{\omega})} : \mathcal{C}_X \to \mathcal{C}_{X_{\omega}}$. We define $P_{\omega} = \ker \rho_{(\pi_{\omega}, t_{\omega})}$.

**Lemma 8.4.** For an O-pair $\omega$, the ideal $P_{\omega}$ of $\mathcal{C}_X$ is gauge-invariant and satisfies $\omega_{P_{\omega}} = \omega$.

**Proof.** Clear by the definitions. □

**Proposition 8.5.** For a gauge-invariant ideal $P$ of $\mathcal{C}_X$, we have $P = P_{\omega_P}$ and $\mathcal{C}_X / P \cong \mathcal{C}_{X_{\omega_P}}$.

**Proof.** If $P$ is gauge-invariant, the representation $(\sigma_P \circ \pi_X, \sigma_P \circ t_X)$ admits a gauge action, where $\sigma_P : \mathcal{C}_X \to \mathcal{C}_X / P$ is a natural surjection. Hence by the definition of $\omega_P$ and Theorem 7.1, we have an isomorphism $\rho : \mathcal{C}_{X_{\omega_P}} \to \mathcal{C}_X / P$ such that $(\rho \circ \pi_{\omega_P}, \rho \circ t_{\omega_P}) = (\sigma_P \circ \pi_X, \sigma_P \circ t_X)$. Hence $\mathcal{C}_X / P \cong \mathcal{C}_{X_{\omega_P}}$ and $P = P_{\omega_P}$. □
Now we get the following.

**Theorem 8.6.** The set of all gauge-invariant ideals of \( \mathcal{C}_X \) corresponds bijectively to the set of all \( O \)-pairs of \( X \) by \( P \mapsto \omega_P \) and \( \omega \mapsto P_\omega \). These maps preserve inclusions and intersections.

In the case that \( C^* \)-correspondences are defined from graphs, or more generally from topological graphs, Theorem 8.6 was proved in [Bates et al. 2002] and [Katsura 2006a].

**Corollary 8.7 [Muhly and Tomforde 2004, Theorem 6.4].** If \( A = J_X + \ker \varphi_X \), then \( P \mapsto I_P \) is a bijection from the set of all gauge-invariant ideals of \( \mathcal{C}_X \) to the set of all invariant ideals of \( A \) with respect to \( X \).

**Proof.** By Theorem 8.6 and Lemma 5.2, it suffices to show that \( J_{X_I} \subset [J_X]_I \) for all invariant ideals \( I \) of \( A \). Let \( I \) be an invariant ideal. Since \( A = J_X + \ker \varphi_X \), we have \( A/I = [J_X]_I + [\ker \varphi_X]_I \). Hence we get \( ([\ker \varphi_X]_I)^{\perp} = [J_X]_I \). Since \( \ker \varphi_X \supset [\ker \varphi_X]_I \), we obtain

\[
J_{X_I} \subset (\ker \varphi_{X_I})^{\perp} \subset ([\ker \varphi_X]_I)^{\perp} = [J_X]_I. \quad \square
\]

Note that the assumption \( A = J_X + \ker \varphi_X \) is equivalent to the assumption in [Muhly and Tomforde 2004, Theorem 6.4]. This is also equivalent to saying that \( A \cong A_1 \oplus A_2 \) and \( \varphi_X: A \to \mathcal{L}(X) \) is the composition of the natural surjection \( A \to A_1 \) and an embedding \( A_1 \hookrightarrow \mathcal{L}(X) \). This assumption is not necessary for the map \( P \mapsto I_P \) to be bijective, as we will see in Sections 9 and 10.

We finish this section with the following result on the gauge-invariant ideals of \( \mathcal{F}_X \).

**Proposition 8.8.** The set of all gauge-invariant ideals of \( \mathcal{F}_X \) corresponds bijectively to the set of all \( T \)-pairs of \( X \) such that inclusions and intersections are preserved.

**Proof.** The set of all gauge-invariant ideals of \( \mathcal{F}_X \) corresponds bijectively to the “set” of all representations of \( X \) admitting gauge actions if we consider two representations \( (\pi, t) \) and \( (\pi', t') \) to be the same when there exists a (necessarily unique) isomorphism \( \rho: C^*(\pi, t) \to C^*(\pi', t') \) such that \( \rho \circ \pi = \pi' \) and \( \rho \circ t = t' \). Under this identification, the “set” of all representations of \( X \) admitting gauge actions corresponds bijectively to the set of all \( T \)-pairs of \( X \) by \( (\pi, t) \mapsto \omega_{(\pi, t)} \) defined in Definition 5.9, and \( \omega \mapsto (\pi_\omega, t_\omega) \) defined in Definition 6.10 by Proposition 6.12 and Theorem 7.1. This completes the proof. \( \square \)


In this section, we prove that each gauge-invariant ideal \( P \) of the \( C^* \)-algebra \( \mathcal{C}_X \) is strongly Morita equivalent to the \( C^* \)-algebra \( \mathcal{C}_Y \) for a certain \( C^* \)-correspondence...
For a positively invariant ideal \(I\) of \(A\), we have \(\varphi_X(I)X \subset XI\). Hence the closed linear subspace \(Y_I = \varphi_X(I)X\) of \(X\) is naturally considered as a \(C^*\)-correspondence over \(I\).

**Lemma 9.1.** For a positively invariant ideal \(I\) of \(A\), we have \(\ker \varphi_Y = I \cap \ker \varphi_X\) and \(\varphi_Y^{-1}(\mathcal{H}(Y)) = I \cap \varphi_X^{-1}(\mathcal{H}(X))\).

**Proof.** Take \(a \in \ker \varphi_Y\). For \(\xi \in X\), we have \(\varphi_X(a)\varphi_X(a^*)\xi = 0\) because \(\varphi_X(a^*)\xi \in Y_I\). Hence we have \(aa^* \in \ker \varphi_X\). Thus we get \(a \in I \cap \ker \varphi_X\). This shows \(\ker \varphi_Y \subseteq I \cap \ker \varphi_X\). Since the converse inclusion is obvious, we get \(\ker \varphi_Y = I \cap \ker \varphi_X\).

Take \(a \in \varphi_Y^{-1}(\mathcal{H}(Y))\). Set \(k = \varphi_Y(a) \in \mathcal{H}(Y) \subset \mathcal{H}(X)\). Since we have \(\varphi_X(a)\varphi_X(b)\xi = k\varphi_X(b)\xi\) for \(b \in I\) and \(\xi \in X\), we get \(\varphi_X(a)\varphi_X(a^*) = k\varphi_X(a)^*\). We also get \(\varphi_X(a)k = k\varphi_X(a)^*\) because \(k \in \mathcal{H}(Y)\). Thus \((\varphi_X(a) - k)\varphi_X(a)^* = 0\). Hence \(\varphi_X(a) = k \in \mathcal{H}(X)\). Since the converse inclusion is obvious, we have \(\varphi_Y^{-1}(\mathcal{H}(Y)) = I \cap \varphi_X^{-1}(\mathcal{H}(X))\).

**Proposition 9.2.** For a positively invariant ideal \(I\) of \(A\), we have \(J_{Y_I} = I \cap J_X\).

**Proof.** Since \(\ker \varphi_Y \subseteq \ker \varphi_X\), we have \((\ker \varphi_Y)^\perp \supset (\ker \varphi_X)^\perp\). By Lemma 9.1, we have \((\ker \varphi_Y)^\perp \cap I \cap \ker \varphi_X = 0\). Hence \((\ker \varphi_Y)^\perp \cap I \subset (\ker \varphi_X)^\perp\). Thus we have \((\ker \varphi_Y)^\perp \cap I = (\ker \varphi_X)^\perp \cap I\). From this equality and Lemma 9.1, we get

\[
J_{Y_I} = \varphi_Y^{-1}(\mathcal{H}(Y_I)) \cap (\ker \varphi_Y)^\perp \\
= I \cap \varphi_X^{-1}(\mathcal{H}(X)) \cap (\ker \varphi_Y)^\perp \\
= I \cap \varphi_X^{-1}(\mathcal{H}(X)) \cap (\ker \varphi_X)^\perp \\
= I \cap J_X.
\]

**Proposition 9.3.** For a positively invariant ideal \(I\) of \(A\), the \(C^*\)-subalgebra generated by \(\pi_X(I)\) and \(t_X(Y_I)\) is isomorphic to \(\mathcal{C}_Y\), and it is the smallest hereditary \(C^*\)-subalgebra in \(\mathcal{C}_X\) containing \(\pi_X(I)\).

**Proof.** Let \(B\) be the \(C^*\)-subalgebra of \(\mathcal{C}_X\) generated by \(\pi_X(I)\) and \(t_X(Y_I)\). Clearly the restrictions of \(\pi_X\) and \(t_X\) to \(I\) and \(Y_I\) give an injective representation \((\pi, t)\) of \(Y_I\) on \(\mathcal{C}_X\) which admits a gauge action. It is also clear that \(C^*(\pi, t) = B\). By Proposition 9.2, this representation \((\pi, t)\) is covariant. Thus \(B\) is isomorphic to \(\mathcal{C}_Y\) by Theorem 3.6.

Since we have \(\pi_X(I)t_X(X)\pi_X(I) = t_X(Y_I)\pi_X(I) = t_X(Y_I)\), \(B\) is contained in the \(C^*\)-subalgebra \(\pi_X(I)\mathcal{C}_X\pi_X(I)\). By [Katsura 2004b, Proposition 2.7], \(\mathcal{C}_X\) is the closure of the linear span of elements in the form

\[
t_X(\xi_1) \cdots t_X(\xi_n) \pi_X(a) t_X(\eta_m)^* \cdots t_X(\eta_1)^*
\]
for \( a \in A \) and \( \xi_k, \eta_l \in X \). Using the fact that \( \pi_X(I) t_X(X) = t_X(Y_I) \pi_X(I) \), we can prove by induction on \( n \) that \( \pi_X(b) t_X(\xi_1) \cdots t_X(\xi_n) \pi_X(a) \in B \) for \( b \in I, a \in A \) and \( \xi_k \in X \). Hence \( \pi_X(I) \mathcal{O}_X \pi_X(I) \) is contained in \( B \). Thus we have shown that \( B = \pi_X(I) \mathcal{O}_X \pi_X(I) \) which is the smallest hereditary \( C^* \)-subalgebra containing \( \pi_X(I) \).

\[ \square \]

**Proposition 9.4.** For an ideal \( I \) of \( A \), the ideal of \( \mathcal{O}_X \) generated by \( \pi_X(I) \) is \( P_\omega \) where \( \omega = (X_{-\infty}(I), X_{-\infty}(I) + J_X) \).

**Proof.** Let \( P \) be the ideal of \( \mathcal{O}_X \) generated by \( \pi_X(I) \). Since \( I \subset I_P \) and \( I_P \) is invariant, we have \( X_{-\infty}(I) \subset I_P \) by Proposition 4.16. Hence we have \( X_{-\infty}(I) + J_X \subset I_P \subset I_P \). Therefore we get \( \omega \subset (I_P, I_P') = \omega_P \). Since \( \pi_X(I) \subset \pi_X(X_{\infty}(I)) \subset P_\omega \), implies \( P \subset P_\omega \), we have \( \omega_P \subset \omega P_\omega = \omega \). Thus we get \( \omega P = \omega \).

Since \( \pi_X(I) \) is closed under the gauge action, the ideal \( P \) is gauge-invariant. Hence we have \( P = P_\omega = P_\omega \) by Proposition 8.5.

\[ \square \]

**Proposition 9.5.** Let \( I \) be a positively invariant ideal of \( A \). For an \( O \)-pair \( \omega = (X_{-\infty}(I), X_{-\infty}(I) + J_X) \), the gauge-invariant ideal \( P_\omega \) is strongly Morita equivalent to the \( C^* \)-algebra \( \mathcal{O}_Y \).

**Proof.** By Proposition 9.3 and Proposition 9.4, the \( C^* \)-subalgebra generated by \( \pi_X(I) \) and \( t_X(Y_I) \) is isomorphic to \( \mathcal{O}_Y \) and is a hereditary and full \( C^* \)-subalgebra of \( P_\omega \) which is the ideal generated by \( \pi_X(I) \). Thus \( P_\omega \) is strongly Morita equivalent to the \( C^* \)-algebra \( \mathcal{O}_Y \).

\[ \square \]

**Corollary 9.6.** Let \( X \) be a \( C^* \)-correspondence over a \( C^* \)-algebra \( A \). Define a \( C^* \)-correspondence \( Y \) over \( A \) by \( Y = \varphi_X(A)X \). Then \( \mathcal{O}_Y \) is strongly Morita equivalent to \( \mathcal{O}_X \).

**Proof.** Apply Proposition 9.5 to the invariant ideal \( A \).

The \( C^* \)-correspondence \( Y \) defined in the above corollary is nondegenerate; that is, \( \varphi_Y(A)Y = Y \). Thus, by Corollary 9.6, we can exchange a given \( C^* \)-correspondence to a nondegenerate one so that the \( C^* \)-algebras constructed by them are strongly Morita equivalent (we used this fact in [Katsura 2004b, Appendix C]).

By Proposition 9.5, gauge-invariant ideals \( P \) satisfying that \( I_P = I_P + J_X \) are shown to be strongly Morita equivalent to the \( C^* \)-algebra \( \mathcal{O}_{Y_{I_P}} \) of the \( C^* \)-correspondence \( Y_{I_P} \). To deal with all gauge-invariant ideals of \( \mathcal{O}(X) \), we need the following argument.

Let us define a \( C^* \)-algebra \( \tilde{A} \) and a Banach space \( \tilde{X} \) by

\[
\tilde{A} = \pi_X(A) + \psi_{t_X}(\mathfrak{H}(X)) \subset \mathcal{O}_X, \\
\tilde{X} = \text{span} \left( t_X(X) + t_X(X) \psi_{t_X}(\mathfrak{H}(X)) \right) \subset \mathcal{O}_X.
\]

If we define the left and right actions of \( \tilde{A} \) on \( \tilde{X} \) as multiplication, and the inner product by \( \langle \xi, \eta \rangle_{\tilde{X}} = \xi^* \eta, \tilde{X} \) becomes a \( C^* \)-correspondence over \( \tilde{A} \). Since the
embeddings $\widetilde{A} \hookrightarrow \mathcal{O}_X$ and $\widetilde{X} \hookrightarrow \mathcal{O}_X$ give an injective representation of $\widetilde{X}$, we have an injective \ast\text{-}homomorphism from $\mathcal{X}(\widetilde{X})$ onto $\text{span}(\widetilde{X}^\ast)$ $\subset \mathcal{O}_X$. Thus we can identify $\mathcal{X}(\widetilde{X})$ with $\text{span}(\widetilde{X}^\ast)$.

**Lemma 9.7.** We have $J_{\widetilde{X}} = \psi_{t_x}(\mathcal{X}(X)) \subset \widetilde{A}$.

*Proof.* By the identification above, the restriction of $\varphi_{\widetilde{X}}$ to the ideal $\psi_{t_x}(\mathcal{X}(X))$ of $\widetilde{A}$ is just the embedding $\psi_{t_x}(\mathcal{X}(X)) \hookrightarrow \mathcal{X}(\widetilde{X})$. Hence we have $\psi_{t_x}(\mathcal{X}(X)) \subset J_{\widetilde{X}}$. We will prove the converse inclusion. Take $\pi_X(a) + \psi_{t_x}(k) \in J_{\widetilde{X}}$. Then we have $\pi_X(a) \in J_{\widetilde{X}}$. Let $\{u_\lambda\}$ be an approximate unit of $\psi_{t_x}(\mathcal{X}(X))$. It is not difficult to see that $\{\varphi_{\widetilde{X}}(u_\lambda)\}$ is an approximate unit of $\mathcal{X}(\widetilde{X})$ (see [Katsura 2004b, Lemma 5.10]). Since $\varphi_{\widetilde{X}}(\pi_X(a)) \in \mathcal{X}(\widetilde{X})$, we have

$$\varphi_{\widetilde{X}}(\pi_X(a)) = \lim_\lambda \varphi_{\widetilde{X}}(\pi_X(a)) \varphi_{\widetilde{X}}(u_\lambda) = \lim_\lambda \varphi_{\widetilde{X}}(\pi_X(a)u_\lambda) \in \varphi_{\widetilde{X}}(\psi_{t_x}(\mathcal{X}(X))).$$

Hence there exists $k \in \mathcal{X}(X)$ with $\varphi_{\widetilde{X}}(\pi_X(a)) = \varphi_{\widetilde{X}}(\psi_{t_x}(k))$. Therefore we have

$$t_x(\varphi_X(a)\xi) = \pi_X(a)t_x(\xi) = \varphi_{\widetilde{X}}(\pi_X(a))t_x(\xi) = \varphi_{\widetilde{X}}(\psi_{t_x}(k))t_x(\xi) = \psi_{t_x}(k)t_x(\xi) = t_x(k\xi)$$

for each $\xi \in X$. Hence we obtain $\varphi_X(a) = k \in \mathcal{X}(X)$. For $b \in \ker \varphi_X$ we have $\pi_X(b) \in \ker \varphi_{\widetilde{X}}$. Therefore we get $\pi_X(ab) = 0$ for all $b \in \ker \varphi_X$. Thus $a \in \varphi_{\widetilde{X}}^{-1}(\mathcal{X}(X)) \cap (\ker \varphi_X)^\perp = J_{\widetilde{X}}$. Therefore $\pi_X(a) + \psi_{t_x}(k) = \psi_{t_x}(\varphi_X(a) + k) \in \psi_{t_x}(\mathcal{X}(X))$. This shows $J_{\widetilde{X}} \subset \psi_{t_x}(\mathcal{X}(X))$. Thus we get $J_{\widetilde{X}} = \psi_{t_x}(\mathcal{X}(X))$. \hfill $\square$

**Proposition 9.8.** The natural inclusions $\widetilde{A} \hookrightarrow \mathcal{O}_X$ and $\widetilde{X} \hookrightarrow \mathcal{O}_X$ induce an isomorphism $\mathcal{O}_X \cong \mathcal{O}_X$.

*Proof.* It is clear that the pair $(\pi, t)$ of the inclusions $\pi : \widetilde{A} \hookrightarrow \mathcal{O}_X$ and $t : \widetilde{X} \hookrightarrow \mathcal{O}_X$ is an injective representation of $\widetilde{X}$ admitting a gauge action and satisfying $C^\ast(\pi, t) = \mathcal{O}_X$. By Lemma 9.7, the representation $(\pi, t)$ is covariant. Hence we have an isomorphism $\rho_{(\pi, t)} : \mathcal{O}_X \rightarrow \mathcal{O}_X$ by Theorem 3.6. \hfill $\square$

**Proposition 9.9.** For a gauge-invariant ideal $P$ of $\mathcal{O}_X$, we set $\widetilde{I} = \widetilde{A} \cap P$. Then $P$ is strongly Morita equivalent to the $C^\ast$-algebra $\mathcal{O}_{\widetilde{Y}}$ where $\widetilde{Y} = \varphi_{\widetilde{X}}(\widetilde{I})\widetilde{X}$ is a $C^\ast$-correspondence over $\widetilde{I}$.

*Proof.* Since $\widetilde{I}$ is the intersection of $\widetilde{A}$ and the ideal $P$ of $\mathcal{O}_X = \mathcal{O}_X$, the ideal $\widetilde{I}$ is an invariant ideal of $\widetilde{A}$. Let $\widetilde{P}$ be the ideal in $\mathcal{O}_X = \mathcal{O}_X$ generated by $\widetilde{I}$. By Proposition 9.5, $\widetilde{P}$ is strongly Morita equivalent to the $C^\ast$-algebra $\mathcal{O}_{\widetilde{Y}}$. We will show that $\widetilde{P} = P$. To do so, it suffices to see $\omega_{\widetilde{P}} = \omega_P$ by Theorem 8.6 because both $\widetilde{P}$ and $P$ are gauge-invariant. Since $\widetilde{I} \subset P$, we have $\widetilde{P} \subset P$. Hence $\omega_{\widetilde{P}} \subset \omega_P$. We have

$$\pi_X(A) \cap P = \pi_X(A) \cap \widetilde{A} \cap P = \pi_X(A) \cap \widetilde{I} \subset \pi_X(A) \cap \widetilde{P}.$$
Similarly,
\[
\pi_X(A) \cap \left( P + \psi_{1x}(\mathcal{K}(X)) \right) = \pi_X(A) \cap \left( \tilde{A} \cap P + \psi_{1x}(\mathcal{K}(X)) \right)
\]
\[
= \pi_X(A) \cap \left( \tilde{I} + \psi_{1x}(\mathcal{K}(X)) \right)
\]
\[
\subset \pi_X(A) \cap \left( \tilde{P} + \psi_{1x}(\mathcal{K}(X)) \right).
\]

Hence we get \( \omega_p \subset \omega_{\tilde{p}} \). Thus \( \omega_{\tilde{p}} = \omega_p \).

**Remark 9.10.** As we saw in the proof of Proposition 9.9, we can see that gauge-invariant ideals of \( \mathcal{O}_X \) are distinguished by their intersection with \( \tilde{A} \). By Proposition 9.9, the set of all gauge-invariant ideals of \( \mathcal{O}_X \) invariance of \( \tilde{A} \) even though the \( C^* \)-correspondence \( \tilde{X} \) does not satisfy the assumption in Corollary 8.7 in general.

Proposition 9.9 shows that every gauge-invariant ideal of \( \mathcal{O}_X \) is strongly Morita equivalent to the \( C^* \)-algebra \( \mathcal{O}_Y \) for some \( C^* \)-correspondences \( Y \). In the next section, we will see that for every gauge-invariant ideal \( P \) of \( \mathcal{O}_X \) we can find a \( C^* \)-correspondence \( Y \) so that \( P \) is isomorphic to \( \mathcal{O}_Y \).

**10. Crossed products by Hilbert \( C^* \)-bimodules**

For a \( C^* \)-algebra \( A \), a **Hilbert \( A \)-bimodule** is a \( C^* \)-correspondence \( X \) over \( A \) together with a left inner product \( \langle \cdot, \cdot \rangle : X \times X \to A \) such that \( \varphi_X(\langle \xi, \eta \rangle) = \theta_{\xi, \eta} \) for \( \xi, \eta \in X \) (for details, see [Abadie et al. 1998], for example). We have
\[
J_X = \overline{\text{span}}\{ \langle \xi, \eta \rangle \in A \mid \xi, \eta \in X \}.
\]

A \( C^* \)-correspondence \( X \) has a left inner product so that it becomes a Hilbert \( A \)-bimodules if and only if we have \( \varphi_X(J_X) = \mathcal{K}(X) \), and in this case a left inner product is uniquely determined by the structure of \( C^* \)-correspondence as \( \langle \xi, \eta \rangle = (\varphi_X|_{J_X})^{-1}(\theta_{\xi, \eta}) \) in \( J_X \) (see [Katsura 2003a, Lemma 3.4]).

For a general \( C^* \)-correspondence \( X \) over \( A \), an ideal \( I \) of \( A \) is positively invariant if and only if \( \varphi_X(I)X \subset XI \). For Hilbert \( C^* \)-bimodules, we get an analogous statement for negative invariance. Let us fix a Hilbert \( A \)-bimodule \( X \) whose left inner product is denoted by \( \langle \cdot, \cdot \rangle \).

**Lemma 10.1.** An ideal \( I \) of \( A \) is negatively invariant if and only if \( \varphi_X(I)X \supset XI \).

**Proof.** Let \( I \) be a negatively invariant ideal of \( A \). Take \( \xi \in X \) and \( a \in I \). For arbitrary \( \eta \in X \), we have \( \varphi_X(\langle \xi a, \eta \rangle) = \theta_{\xi a, \eta} \in \mathcal{K}(XI) \). Since \( \langle \xi a, \eta \rangle \in J_X \), the negative invariance of \( I \) implies \( \varphi_X(\langle \xi \eta \rangle) \in I \) for arbitrary \( \eta \in X \). Similarly to the proof of Proposition 1.3, we can prove \( \xi a \in \varphi_X(I)X \). Thus we have \( \varphi_X(I)X \supset XI \). Conversely, assume that an ideal \( I \) satisfies \( \varphi_X(I)X \supset XI \). For \( \xi, \eta \in XI \), we can find \( \xi' \in X \) and \( a \in I \) with \( \xi = \varphi_X(a)\xi' \). Therefore we have \( \varphi_X(a)\xi' = \varphi_X(a)\xi' \) for arbitrary \( \eta \in X \).
\( \langle \phi_X(a)\xi', \eta \rangle = a(\xi', \eta) \in I \). Hence we can see that \((\varphi_X|_I)^{-1}(k) \in I \) for \( k \in \mathcal{K}(XI) \). Therefore for \( a \in J_X \) with \( \varphi_X(a) \in \mathcal{K}(XI) \) we have \( a \in I \). This shows that \( I \) is negatively invariant. \( \square \)

**Proposition 10.2.** An ideal \( I \) of \( A \) is invariant if and only if \( \varphi_X(I)X = XI \).

**Proof.** Clear from Lemma 10.1. \( \square \)

**Proposition 10.3.** For an invariant ideal \( I \) of \( A \), the C*-correspondence \( X_I \) defined in Section 5 has a left inner product \( X_I \langle \cdot, \cdot \rangle \) such that \( X_I \langle [\xi], [\eta] \rangle = [X(\langle \xi, \eta \rangle)]I \) for \( \xi, \eta \in X \).

**Proof.** Since \( \varphi_X(I)X = XI \), it is not difficult to see that the left inner product of \( X_I \) described above is well-defined, and satisfies the required conditions. \( \square \)

**Corollary 10.4.** For an invariant ideal \( I \) of \( A \), we have \( J_X = [J_X]I \).

**Proof.** By Proposition 10.3, we have

\[
J_X = \overline{\text{span}} \{X_I \langle \xi', \eta \rangle \in A/I \ | \ \xi', \eta \in X_I \} = \overline{\text{span}} \{[X(\langle \xi, \eta \rangle)]I \in A/I \ | \ \xi, \eta \in X \} = [J_X]I.
\]

\( \square \)

**Proposition 10.5.** For an invariant ideal \( I \), the C*-subalgebra of \( \mathcal{O}_X \) generated by \( \pi_X(I) \) and \( t_X(X_I) \) is an ideal.

**Proof.** This follows from the fact that \( XI = \varphi_X(I)X = \varphi_X(I)XI \). \( \square \)

**Theorem 10.6.** Let \( X \) be a Hilbert \( A \)-bimodule. For an ideal \( P \) of \( \mathcal{O}_X \), we define an ideal \( I_P \) of \( A \) by \( \pi_X(I_P) = \pi_X(A) \cap P \). Then the map \( P \mapsto I_P \) gives a one-to-one correspondence from the set of all gauge-invariant ideals \( P \) of \( \mathcal{O}_X \) to the set of ideals \( I \) of \( A \) satisfying \( \varphi_X(I)X = XI \). We also have isomorphisms \( P \cong \mathcal{O}_{XI_P} \) and \( \mathcal{O}_X/P \cong \mathcal{O}_{XI_P} \) for a gauge-invariant ideal \( P \).

**Proof.** By Corollary 10.4, we have \( I' = I + J_X \) for all \( O \)-pair \( \omega = (I, I') \). Thus the first assertion follows from Theorem 8.6 and Proposition 10.2. The second assertion follows from Proposition 8.5, Proposition 9.3 and Proposition 10.5. \( \square \)

Both \( XI \) and \( X_I \) are Hilbert C*-bimodules. Thus the class of C*-algebras associated with Hilbert C*-bimodules behave well. We will see that this class is same as the one of C*-algebras associated with C*-correspondences, which we are studying in this paper.

Let us take a C*-algebra \( A \) and a C*-correspondence \( X \) over \( A \). We define a C*-algebra \( \bar{A} \) and a Banach space \( \bar{X} \) by

\[
\bar{A} = \mathcal{O}_{X'}, \quad \bar{X} = \{ x \in \mathcal{O}_X \ | \ \gamma_z(x) = zx \text{ for all } z \in T \}.
\]
Remark 10.7. In a similar way as in the proof of [Katsura 2004b, Proposition 5.7], we can prove that $\mathcal{X} = \overline{\text{span}}(t_X(X)\mathcal{O}_X^\gamma)$. We do not use this fact.

It is easy to see that $\mathcal{X}$ is a Hilbert $\mathcal{A}$-bimodule where the inner products are defined by
$$\langle \xi, \eta \rangle = \xi^* \eta, \quad \mathcal{X} \langle \xi, \eta \rangle = \xi \eta^*,$$
for $\xi, \eta \in \mathcal{X}$, and the left and right actions are multiplication.

Proposition 10.8 (compare [Abadie et al. 1998, Theorem 3.1]). The natural embedding of $\mathcal{A}$ and $\mathcal{X}$ into $\mathcal{O}_X$ gives an isomorphism $\mathcal{O}_\mathcal{X} \cong \mathcal{O}_X$.

Proof. By Theorem 3.6, it suffices to check that the embedding of $\mathcal{A}$ and $\mathcal{X}$ into $\mathcal{O}_X$ is an injective covariant representation admitting a gauge action. These conditions are easily checked. □

Corollary 10.9. Let $X$ be a $C^*$-correspondence over a $C^*$-algebra $A$, and $P$ be a gauge-invariant ideal of $\mathcal{O}_X$. If we set $I = P \cap A$, then $P$ is isomorphic to $\mathcal{O}_{\mathcal{X} I}$.


We remark that in order to compute the $K$-groups of gauge-invariant ideals, Proposition 9.5 and Proposition 9.9 seem to be more useful than Corollary 10.9.

11. Relative Cuntz–Pimsner algebras

In this last section, we apply the results obtained above to the relative Cuntz–Pimsner algebras introduced in [Muhly and Solel 1998]. Recall that for a $C^*$-correspondence $X$ over a $C^*$-algebra $A$, and an ideal $J$ of $A$ with $\varphi_X(J) \subset \mathcal{H}(X)$, the relative Cuntz–Pimsner algebra $\mathcal{O}(J, X)$ is generated by the image of a representation $(\pi, t)$ which is universal among representations satisfying $\pi(a) = \psi_t(\varphi_X(a))$ for $a \in J$ (see [Muhly and Solel 1998, Theorem 2.19]). We will show that every relative Cuntz–Pimsner algebra is isomorphic to $\mathcal{O}_{\mathcal{X} I}$ for some $C^*$-correspondences $X'$. In particular, every Cuntz–Pimsner algebra and Toeplitz algebra introduced in [Pimsner 1997], including augmented ones, is in the class of our $C^*$-algebras.

By universality, the representation $(\pi, t)$ of $X$ on $\mathcal{O}(J, X)$ admits a gauge action. Hence by Corollary 7.13 we see that $\mathcal{O}(J, X)$ is isomorphic to $\mathcal{O}_{\mathcal{X}_{\sigma(\pi, t)}}$. We will express $\sigma(\pi, t)$ in terms of a $C^*$-correspondence $X$ over $A$ and an ideal $J$ of $A$.

Now let us take a $C^*$-correspondence $X$ over a $C^*$-algebra $A$, and an ideal $J$ of $A$ with $\varphi_X(J) \subset \mathcal{H}(X)$. We inductively define an increasing family of ideals $\{J_n\}_{n \in \mathbb{N}}$ by $J_0 = 0$ and $J_{n+1} = J_n + J \cap X^{-1}(J_{n+1})$. We set $J_{-\infty} = \lim_{n \to \infty} J_n$. We denote by $\omega_J$ the pair $(J_{-\infty}, J)$ of ideals of $A$. Since $X^{-1}(0) = \ker \varphi_X$, we have $J_{-1} = J \cap \ker \varphi_X$. It is easy to see that $J_{-\infty} = 0$ if and only if $J \cap \ker \varphi_X = 0$.

Lemma 11.1. The pair $\omega_J = (J_{-\infty}, J)$ is a $T$-pair of $X$. 

Proof: Clearly $J_0 = 0$ is positively invariant. We can prove that $J_{-n}$ is positively invariant for all $n \in \mathbb{N}$ by induction with respect to $\mathbb{N}$, as in Lemma 4.15. Hence $J_{-\infty}$ is positively invariant. Again by induction, we see that $J_{-\infty} \subset J$. Since $J \cap X^{-1}(J_{-n}) \subset J_{-(n+1)} \subset J_{-\infty}$ for all $n$, we have $J \cap X^{-1}(J_{-\infty}) \subset J_{-\infty}$ by Proposition 4.7. Since $\varphi_X(J) \subset \mathcal{H}(X)$ by assumption, we have $[\varphi_X(J)]_{-\infty} \subset \mathcal{H}(X_{J_{-\infty}})$. Hence we get $J \subset J(J_{-\infty})$. Thus we have $J_{-\infty} \subset J \subset J(J_{-\infty})$. We are done.

Lemma 11.2. If a $T$-pair $\omega = (I, I')$ satisfies $J \subset I'$, then $\omega_J \subset \omega$.

Proof. We will prove $J_{-n} \subset I$ by induction on $n$. For $n = 0$, it is trivial. Assume $J_{-n} \subset I$. We have

$$J \cap X^{-1}(J_{-n}) \subset I' \cap X^{-1}(I) \subset J(I) \cap X^{-1}(I) = I$$

by Lemma 5.2. Hence

$$J_{-(n+1)} = J_{-n} + J \cap X^{-1}(J_{-n}) \subset I.$$ 

We have shown that $J_{-n} \subset I$ for all $n$. This implies that $J_{-\infty} \subset I$. Hence we have $\omega_J \subset \omega$.

Proposition 11.3. The relative Cuntz–Pimsner algebra $\mathcal{O}(J, X)$ is isomorphic to the $C^*$-algebra $\mathcal{O}_{X_{\omega_J}}$ of the $C^*$-correspondence $X_{\omega_J}$.

Proof. Let us denote by $(\pi, t)$ the universal representation of $X$ on $\mathcal{O}(J, X)$ satisfying $\pi(a) = \psi_t(\varphi_X(a))$ for all $a \in J$, and by $(\pi_{\omega_J}, t_{\omega_J})$ the representation of $X$ on the $C^*$-algebra $\mathcal{O}_{X_{\omega_J}}$ defined in Section 6. By Proposition 6.12, we have $I'_{(\pi_{\omega_J}, t_{\omega_J})} = J$. Hence by Lemma 5.10 (v), we have $\pi_{\omega_J}(a) = \psi_{t_{\omega_J}}(\varphi_X(a))$ for all $a \in J$. By the universal property of $\mathcal{O}(J, X)$, there exists a $*$-homomorphism $\rho: \mathcal{O}(J, X) \to \mathcal{O}_{X_{\omega_J}}$ such that $\pi_{\omega_J} = \rho \circ \pi$ and $t_{\omega_J} = \rho \circ t$. On the other hand, $J \subset I'_{(\pi, t)}$ implies $\omega_J \subset \omega_{(\pi, t)}$ by Lemma 11.2. Hence by Theorem 7.1, there exists a surjective $*$-homomorphism $\rho': \mathcal{O}_{X_{\omega_J}} \to \mathcal{O}(J, X)$ such that $\pi = \rho' \circ \pi_{\omega_J}$ and $t = \rho' \circ t_{\omega_J}$. Clearly $\rho$ and $\rho'$ are the inverses of each others. Hence $\mathcal{O}(J, X)$ is isomorphic to $\mathcal{O}_{X_{\omega_J}}$.

By Proposition 11.3, the $T$-pair $\omega_{(\pi, t)}$ arising from the representation $(\pi, t)$ on $\mathcal{O}(J, X)$ coincides with $\omega_J = (J_{-\infty}, J)$. From this fact, we have the following corollaries.

Corollary 11.4. Let $(\pi, t)$ be the representation of $X$ on $\mathcal{O}(J, X)$. Then the kernel of the map $\pi: A \to \mathcal{O}(J, X)$ is $J_{-\infty}$, and we have

$$\{a \in A \mid \varphi_X(a) \in \mathcal{H}(X), \text{ and } \pi(a) = \psi_t(\varphi_X(a))\} = J.$$ 

Proof. This easily follows from $\omega_{(\pi, t)} = \omega_J$. 

Corollary 11.5. The relative Cuntz–Pimsner algebra \( \mathcal{O}(J, X) \) is zero if and only if \( J_{-\infty} = A \).

Proof. Clear by Proposition 11.3.

Corollary 11.6 [Muhly and Solel 1998, Proposition 2.21]. The map \( \pi : A \rightarrow \mathcal{O}(J, X) \) is injective if and only if \( J \cap \ker \varphi_X = 0 \).

Proof. By Corollary 11.4, \( \pi : A \rightarrow \mathcal{O}(J, X) \) is injective if and only if \( J_{-\infty} = 0 \), which is equivalent to the condition \( J \cap \ker \varphi_X = 0 \) as we saw above.

We now state a gauge-invariant uniqueness theorem for relative Cuntz–Pimsner algebras.

Corollary 11.7. For a representation \( (\pi', t') \) of \( X \) satisfying \( \pi'(a) = \psi_{t'}(\varphi_X(a)) \) for \( a \in J \), the natural surjection \( \mathcal{O}(J, X) \rightarrow C^*(\pi', t') \) is an isomorphism if and only if \( (\pi', t') \) admits a gauge action, \( \ker \pi' = J_{-\infty} \), and

\[
\{ a \in A \mid \pi'(a) \in \psi_{t'}(\mathcal{H}(X)) \} = J.
\]

Proof. By Proposition 11.3, \( \mathcal{O}(J, X) \) is canonically isomorphic to \( \mathcal{O}_{X_{\omega J}} \). By Theorem 7.1, the surjection from \( \mathcal{O}_{X_{\omega J}} \) to \( C^*(\pi', t') \) is injective if and only if \( (\pi', t') \) admits a gauge action and \( \omega_{(\pi', t')} = \omega_J \). The last two conditions in the statement just rephrase the condition \( \omega_{(\pi', t')} = \omega_J \).

Note that we automatically have \( \ker \pi' \supset J_{-\infty} \) and \( \{ a \in A \mid \pi'(a) \in \psi_{t'}(\mathcal{H}(X)) \} \supset J \). Note also that in general we cannot replace the condition

\[
\{ a \in A \mid \pi'(a) \in \psi_{t'}(\mathcal{H}(X)) \} = J.
\]

with the condition

\[
\{ a \in A \mid \varphi_X(a) \in \mathcal{H}(X), \text{ and } \pi'(a) = \psi_{t'}(\varphi_X(a)) \} = J,
\]

which seems natural at first glance. This is because there may exist \( a \in A \) with \( \varphi_X(a) \notin \mathcal{H}(X) \) satisfying \( \{ \varphi_X(a) \}_{J_{-\infty}} \in \mathcal{H}(X_{J_{-\infty}}) \) and \( \pi'(a) = \psi_{t'}([\varphi_X(a)]_{J_{-\infty}}) \in \psi_{t'}(\mathcal{H}(X)) \) (see Lemma 5.10 (iv) and (v)). In the case that \( J \cap \ker \varphi_X = 0 \), the statement of Corollary 11.7 has the following simple forms.

Corollary 11.8. Let us assume \( J \cap \ker \varphi_X = 0 \). For a representation \( (\pi', t') \) of \( X \) satisfying \( \pi'(a) = \psi_{t'}(\varphi_X(a)) \) for \( a \in J \), the natural surjection \( \mathcal{O}(J, X) \rightarrow C^*(\pi', t') \) is an isomorphism if and only if \( (\pi', t') \) is injective, admits a gauge action, and satisfies

\[
\{ a \in A \mid \varphi_X(a) \in \mathcal{H}(X), \text{ and } \pi'(a) = \psi_{t'}(\varphi_X(a)) \} = J.
\]

We remark that an ideal \( J \) of \( A \) satisfies \( \varphi_X(J) \subset \mathcal{H}(X) \) and \( J \cap \ker \varphi_X = 0 \) if and only if \( J \subset J_X \). As we saw in Corollary 11.6, the maps from \( A \) and \( X \) to the relative Cuntz–Pimsner algebra \( \mathcal{O}(J, X) \) is injective only when \( J \) satisfies
Thus it is not a good idea to examine the structure of $\mathcal{O}(J, X)$ in terms of $A$, $X$ and $J$ unless $J$ satisfies $J \subset J_x$. Anyway, the following result on the ideal structure of relative Cuntz–Pimsner algebras $\mathcal{O}(J, X)$ can be easily obtained similarly as Theorem 8.6 or Proposition 8.8.

**Proposition 11.9.** Let $X$ be a $C^*$-correspondence over a $C^*$-algebra $A$, and $J$ be an ideal of $A$ with $\varphi_X(J) \subset \mathcal{B}(X)$. Then there exists a one-to-one correspondence between the set of all gauge-invariant ideals of $\mathcal{O}(J, X)$ and the set of all $T$-pairs $\omega = (I, I')$ of $X$ satisfying $J \subset I'$, which preserves inclusions and intersections.

We note that a $T$-pair $\omega = (I, I')$ satisfies $J \subset I'$ if and only if $\omega_J \subset \omega$ by Lemma 11.2.

**Acknowledgements**

The author is grateful to Yasuyuki Kawahigashi for his constant encouragement. This paper was written while the author was staying at the University of Oregon. He would like to thank people there for their warm hospitality.

**References**


Received October 25, 2003.

TAKESHI KATSURA
DEPARTMENT OF MATHEMATICS
HOKKAIDO UNIVERSITY
KITA 10, NISHI 8
KITA-KU, SAPPORO, 060-0810
JAPAN

katsura@math.sci.hokudai.ac.jp
NONDEGENERACY OF COVERINGS OF MINIMAL TORI AND KLEIN BOTTLES IN RIEMANNIAN MANIFOLDS

JOHN DOUGLAS MOORE

We say that a parametrized minimal torus or Klein bottle in an ambient Riemannian manifold is Morse nondegenerate if it lies on a nondegenerate critical submanifold which is also an orbit for the group of isometries of the flat metric of total area one. We show that for a generic choice of a Riemannian metric on a compact manifold of dimension at least four, unbranched multiple covers of prime minimal tori or Klein bottles are Morse nondegenerate. A similar result holds for harmonic tori and Klein bottles. The proofs require a modification of techniques of Bott for studying iterations of smooth closed geodesics.

1. Introduction

Suppose that $\text{Map}(\Sigma, M)$ is a suitable completion of the space of smooth maps $f : \Sigma \to M$ from a compact connected surface $\Sigma$ into a Riemannian manifold $M$ and that $\mathcal{F}$ is the space of marked conformal structures on $\Sigma$. (We complete with respect to a Sobolev norm strong enough so that $\text{Map}(\Sigma, M)$ is a smooth infinite-dimensional Banach manifold and there is a continuous inclusion $\text{Map}(\Sigma, M) \subset C^0(\Sigma, M)$ inducing an isomorphism on all homotopy groups.) A parametrized minimal surface can then be regarded as a critical point for the energy function

$$E : \text{Map}(\Sigma, M) \times \mathcal{F} \to \mathbb{R} \text{ defined by } E(f, \omega) = \frac{1}{2} \int_{\Sigma} |df|^2 dA,$$

where $|df|$ and $dA$ are calculated with respect to some Riemannian metric on $\Sigma$ which lies within the conformal class $\omega \in \mathcal{F}$. A nonconstant parametrized minimal surface $f : \Sigma \to M$ is prime if it is not a nontrivial cover (possibly branched) of a parametrized minimal surface $f_0 : \Sigma_0 \to M$ of lower energy, $\Sigma_0$ being allowed to be nonorientable. In [Moore 2006] we proved a bumpy metric theorem which states that when a compact manifold $M$ of dimension at least four is given a generic Riemannian metric, all prime parametrized minimal surfaces $f : \Sigma \to M$ are free of

MSC2000: primary 53C40, 58E12; secondary 58D15, 58E05.
Keywords: minimal surfaces, harmonic surfaces, bumpy metrics, generic Riemannian metrics.
branch points and are as Morse nondegenerate (as critical points for $E$) as allowed by the group of conformal automorphisms of $\Sigma$.

To be more precise about what we mean by Morse nondegenerate, we first recall that when $\Sigma$ is the sphere (or the projective plane), $E$ is invariant under an action of the six-dimensional Lie group $G = PSL(2, \mathbb{C})$ by linear fractional transformations on the range, while if $\Sigma$ is the torus (or the Klein bottle), $E$ is invariant under an action of the two-dimensional group $G = S^1 \times S^1$. Thus nonconstant minimal spheres or tori must lie on orbits of critical points of dimension six or two, respectively. If $\Sigma$ (or its double cover) has genus at least two, we let $G$ denote the trivial group.

**Definition.** Let $F : \mathcal{M} \to \mathbb{R}$ be a $C^2$ function on a smooth manifold $\mathcal{M}$ which is modeled on a Hilbert or Banach space. A nondegenerate critical submanifold of $\mathcal{M}$ is a finite-dimensional submanifold $N \subset \mathcal{M}$ such that every $f \in N$ is a critical point for $F$, and

$$f \in N \implies T_f N = \{X \in T_f \mathcal{M} : d^2 F(f)(X, Y) = 0 \text{ for all } Y \in T_f \mathcal{M}\}.$$  

Here $d^2 F$ is the Hessian of $F$ at the critical point, and elements $X \in T_f \mathcal{M}$ which satisfy the condition on the right-hand side of (2) are called Jacobi fields.

The notion of a nondegenerate critical submanifold is due to Bott [1982] and plays a large role in the Morse theory of closed geodesics.

**Definition.** We say that a parametrized minimal surface $f : \Sigma \to M$ is Morse nondegenerate if either $f$ is a Morse nondegenerate critical point for $E$ in the usual sense of Morse theory, or $f$ lies on a $G$-orbit which is a nondegenerate critical submanifold for $E$.

With these definitions in place, the bumpy metric theorem of [Moore 2006] can be restated: if $M$ has dimension at least three, then for generic choice of Riemannian metric on $M$, all prime parametrized minimal surfaces are Morse nondegenerate. This can be regarded as an analog of Abraham’s [1970] bumpy metric theorem for smooth closed geodesics, which asserts that for generic metrics on $M$, all smooth closed geodesics lie on nondegenerate critical submanifolds of $\text{Map}(S^1, M)$, each an orbit for the $S^1$ action. However, an important difference is that Abraham’s bumpy metric theorem applies to all closed geodesics, prime or not.

In analogy with the theory of smooth closed geodesics, one might hope that unbranched covers of tori and Klein bottles also lie on nondegenerate critical submanifolds for generic metrics, and this is in fact the case:

**Theorem 1.** Suppose that $M$ is a compact smooth manifold of dimension at least four with Riemannian metric $g_0$. 

(1) If \( f_0 : \Sigma_0 \to M \) is a Morse nondegenerate prime minimal torus or Klein bottle with no branch points, then for a generic choice of Riemannian metric on \( M \) near \( g_0 \), all minimal tori and Klein bottles which cover \( f_0 \) are also Morse nondegenerate.

(2) If \( f_0 : \Sigma \to M \) is a nonoriented Morse nondegenerate prime minimal surface of any genus with no branch points, then for a generic choice of Riemannian metric on \( M \) near \( g_0 \), the oriented double cover of \( f_0 \) is also Morse nondegenerate.

We adopt the convention that the genus of a connected nonorientable surface is the genus of its orientable double cover. From this theorem and the Main Theorem of [Moore 2006], it follows that for generic choice of Riemannian metric on \( M \), all parametrized minimal tori are free of branch points and Morse nondegenerate, except for branched covers of minimal two-spheres, which are of course forced to have branch points. Although it can be shown that branched covers of spheres by tori lie on smooth submanifolds of dimension \( 2d + 2 \), where \( d \) is the degree, there is no reason to suspect that these submanifolds should be Morse nondegenerate for generic choice of Riemannian metric.

Thus if \( M \) is not simply connected, all parametrized minimal tori in one of the nontrivial components of \( \text{Map}(T^2, M) \times \mathcal{T} \) lie on nondegenerate critical submanifolds of dimension two for generic choice of Riemannian metric on \( M \). The motivation behind Theorem 1 is that it serves as part of the foundation necessary for a study of Morse theory for parametrized minimal surfaces via perturbation, using the \( \alpha \)-energy of Sacks and Uhlenbeck [1981; 1982].

Following [McDuff and Salamon 2004], we say that a conformal harmonic map \( f : \Sigma \to M \) from a compact surface \( \Sigma \) is somewhere injective if there exists at least one point \( p \in \Sigma \) such that \( f^{-1}(f(p)) = \{p\} \). It then follows from unique continuation theorems that the set of points \( p \) for which \( f^{-1}(f(p)) = \{p\} \) is open and dense. Note that if \( f \) is an imbedding, it is somewhere injective, and indeed, \( f^{-1}(f(p)) = \{p\} \) for every \( p \in \Sigma \). On the other hand, nontrivial branched covers are not somewhere injective. It follows from the theory of branched immersions developed by Gulliver, Osserman and Royden [Gulliver et al. 1973], or directly from Lemma 4.1 of [Moore 2006], that prime parametrized minimal surfaces are always somewhere injective, and the proof of the Main Theorem of the latter article uses this fact.

To prove that multiple covers are nondegenerate for generic choice of metric, we extend part of Bott’s theory of iterated closed geodesics [Bott 1956]. In the argument, we use the following fact that can be proven using the implicit function theorem. If a prime parametrized minimal surface \( f_0 \) is Morse nondegenerate for a metric \( g_0 \) on \( M \), then there is a unique \( G \)-orbit of Morse nondegenerate minimal
surfaces with the same topology near \( f_0 \) for any metric near \( g_0 \). (If \( \Sigma \) is the two-sphere or projective plane, it is convenient to replace the group \( \text{PSL}(2, \mathbb{C}) \) by its compact subgroup \( \text{SO}(3) \), by fixing the center of mass.) Thus it suffices to consider perturbations in the metric \( g_0 \) which have the same one-jet as \( g_0 \) along \( f_0 \), and therefore preserve minimality of \( f_0 \). A corresponding fact was used in the proof given in [Klingenberg and Takens 1972; Klingenberg 1978] for the bumpy metric theorem for closed geodesics, and indeed our argument can be thought of as an extension of that proof from one to two dimensions.

A similar bumpy metric theorem (Theorem 2) will be formulated for harmonic tori and Klein bottles in Section 5 and proved in Section 6. We will see that in the case of nonconformal harmonic maps, we must deal with a new phenomenon, the possibility of fold points.

2. Preliminaries

We first recall some basic concepts from the theory of harmonic and minimal surfaces in Riemannian manifolds, further details being found in [Micallef and Moore 1988] and [Moore 2006]. If we fix the conformal structure \( \omega \in \mathcal{F} \), the energy reduces to the \( \omega \)-energy

\[ E_\omega : \text{Map}(\Sigma, M) \to \mathbb{R}, \quad \text{defined by} \quad E_\omega(f) = E(f, \omega), \]

the critical points of which are called \( \omega \)-harmonic maps. To determine the equation for harmonic maps, we take the first derivative of \( E_\omega \), obtaining

\[ dE_\omega(f)(X) = \int_\Sigma \langle F_\omega(f, g), X \rangle dA, \quad \text{for} \quad X \in \Gamma(f^*TM), \]

and set \( F_\omega(f, g) = 0 \), where \( F_\omega(\cdot, g) \) is the Euler–Lagrange operator. If one chooses local conformal coordinates \((x_1, x_2)\) on \( \Sigma \), and lets \( \lambda^2 \) denote the conformal factor in the metric, so that the area element is given by \( dA = \lambda^2 dx_1 dx_2 \), the equation one obtains is

\[ \frac{D}{\partial x_1} \left( \frac{\partial f}{\partial x_1} \right) + \frac{D}{\partial x_2} \left( \frac{\partial f}{\partial x_2} \right) = 0, \quad \text{where} \quad \frac{\partial f}{\partial x_a} = f_* \left( \frac{\partial}{\partial x_a} \right) \]

is regarded as a section of the bundle \( f^*TM \) over \( \Sigma \) for \( a = 1, 2 \), and \( D \) denotes the pullback of the Levi-Civita connection of \( M \) to \( f^*TM \).

Differentiating \( E_\omega \) once again gives us the Hessian at a critical point,

\[ d^2E_\omega(f)(X, Y) = \int_\Sigma \langle D_1 F_\omega(f, g)(X), Y \rangle dA = \int_\Sigma \langle L_\omega(X), Y \rangle dA, \]

for \( X, Y \in \Gamma(f^*TM) \). Here \( D_1 F_\omega \) denotes the derivative with respect to the variable \( f \in \text{Map}(\Sigma, M) \) and \( L_\omega \) is the \textit{Jacobi operator} for \( E_\omega \), which acts on sections
of the pullback $f^*TM$ of the tangent bundle to $M$, or equivalently, elements of the tangent space to $\text{Map}(\Sigma, M)$ at $f$. The Jacobi equation is $L_{\omega}(X) = 0$ when $X \in \Gamma(f^*TM)$, and solutions to this equation are called Jacobi fields for the function $E_{\omega}$. A well-known calculation shows that

$$d^2E_{\omega}(f)(X, X) = \int_\Sigma \left( \|DX\|^2 - \langle \mathcal{E}(X), X \rangle \right) dA,$$

where in terms of the complex parameter $z = x_1 + i x_2$ on $\Sigma$,

$$\|DX\|^2 = \frac{1}{\lambda^2} \left( \|DX_{\partial x_1}\|^2 + \|DX_{\partial x_2}\|^2 \right)$$

and

$$\langle \mathcal{E}(X), X \rangle = \frac{1}{\lambda^2} \left( R \left( X, \frac{\partial f}{\partial x_1} \right) \frac{\partial f}{\partial x_1}, X \right) + \left( R \left( X, \frac{\partial f}{\partial x_2} \right) \frac{\partial f}{\partial x_2}, X \right),$$

$R$ being the Riemann–Christoffel curvature tensor of $M$.

A harmonic map $f$ is said to be conformal if it satisfies the conditions

$$\left\langle \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_1} \right\rangle = \left\langle \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_2} \right\rangle \quad \text{and} \quad \left\langle \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right\rangle = 0.$$

Parametrized minimal surfaces, the critical points of the two-variable energy $E$, are exactly the conformal harmonic maps.

The only possible singularities of parametrized minimal surfaces are branch points. These are most easily described in terms of a local complex parameter $z = x_1 + i x_2$ (for the oriented double cover if $\Sigma$ is nonorientable). If we let

$$\frac{\partial f}{\partial z} = \frac{1}{2} \left( \frac{\partial f}{\partial x_1} + i \frac{\partial f}{\partial x_2} \right),$$

a section of the complexified tangent bundle $f^*TM \otimes \mathbb{C}$, a point $p \in \Sigma$ is a branch point if $(\partial f/\partial z)(p) = 0$. If the coordinate $z$ is centered at $p$, we can then write $(\partial f/\partial z) = z^v g$, where $g(p)$ is nonzero, and $v$ is the branching order of $f$ at $p$.

If $\Sigma$ is oriented, the locally defined sections $(\partial f/\partial z)$ generate a line bundle $L$ contained in $f^*TM \otimes \mathbb{C}$, which can be extended smoothly to the branch points. If $\Sigma$ is not orientable, the line bundle $L$ can be defined over the oriented double cover. In either case, the real and imaginary parts of sections of $L$ determine a two-dimensional subbundle $(f^*TM)^\top$ of $f^*TM$, which possesses an orthogonal complement $(f^*TM)^\perp$, yielding a direct sum decomposition

$$f^*TM = (f^*TM)^\top \oplus (f^*TM)^\perp.$$

Thus if $f$ is a conformal harmonic map, we can speak of tangent and normal sections of $f^*TM$ even if $f$ has branch points.
Just as we did for $E_\omega$, we can calculate the first and second derivatives of the two-variable energy function $E : \text{Map}(\Sigma, M) \times \mathcal{T} \rightarrow \mathbb{R}$. These derivatives are calculated in [Moore 2006, §5], and are essential for studying conformal harmonic maps with branch points.

However, the Main Theorem of [Moore 2006] says that for generic choice of metric all parametrized minimal surfaces are free of branch points, and hence immersions. On the space of immersions $f : \Sigma \rightarrow M$, it is simpler to consider the area function $A(f)$, defined by

$$A(f) = \int \left| \frac{\partial f}{\partial x} \wedge \frac{\partial f}{\partial y} \right| \, dx \, dy.$$  

There is a classical formula for second variation of area under normal variations when $f : \Sigma \rightarrow M$ is a minimal immersion, which is presented in [Simons 1968, Theorem 3.2.2], [Lawson 1980, Theorem 32], and many other places. It states that if $X$ is a section of the normal bundle $\Gamma((f^* TM)^\perp)$,

$$d^2 A(f)(X, X) = \int_\Sigma \left( \|DX\|^2 - \langle \mathcal{B}(X) + \mathcal{H}(X), X \rangle \right) \, dA,$$

where $\langle \mathcal{B}(X), X \rangle = \|DX\|^2$.

and $(\cdot)^\perp$ and $(\cdot)^\top$ denote projections into the tangent and normal spaces respectively. It is this second variation formula that we use in the proof of the Theorem. Just as in the case of $E_\omega$, we obtain a formally self-adjoint Jacobi operator

$$L^\perp : \Gamma((f^* TM)^\perp) \rightarrow \Gamma((f^* TM)^\perp),$$

from the second variation formula, such that

$$d^2 A(f)(X_1, X_2) = \int_\Sigma \langle L^\perp(X_1), X_2 \rangle \, dA,$$

for $X_1, X_2 \in \Gamma((f^* TM)^\perp)$, which we call the normal Jacobi operator. We call a solution to $L^\perp(X) = 0$ a normal Jacobi field.

It can be proved that in normal directions (7) gives the same result as second variation of the two-variable energy $E$ when the conformal structure is constrained to move in such a way that conformality of $f$ is preserved.

3. Tori covering tori

When the Riemann surface $\Sigma$ is a torus, the Teichmüller space $\mathcal{T}$ is the upper half plane

$$\mathbb{H} = \{ \omega = u + iv \in \mathbb{C} : v > 0 \},$$

the point $\omega = u + iv$ corresponding to the conformal class of the torus $\mathbb{C}/\Lambda$, where $\Lambda$ is the lattice in $\mathbb{C}$ generated by $1$ and $\omega$. After a change of basis we can arrange
that a given element \( \omega \in \mathcal{T} \) lies in the fundamental domain

\[
D = \left\{ u + iv \in \mathbb{C} : -\frac{1}{2} \leq u \leq \frac{1}{2}, \ u^2 + v^2 \geq 1 \right\}
\]

for the action of the group \( SL(2, \mathbb{Z}) \), the action having kernel \{\pm 1\}. The moduli space \( \mathcal{B} \) for the torus is the quotient of the upper half plane by the \( SL(2, \mathbb{Z}) \)-action, and is obtained from \( D \) by identifying points on the boundary. It is well-known that \( \mathcal{B} \) is diffeomorphic to the space \( \mathbb{C} \) of complex numbers.

The complex torus corresponding to a given \( \omega \in \mathcal{T} \) can be regarded as the quotient of \( \mathbb{C} \) by the abelian subgroup generated by \( d \) and \( od \), where \( d \) is any positive real number, or alternatively, this torus is obtained from a fundamental parallelogram spanned by \( d \) and \( od \) by identifying opposite sides. If \( \omega = u + iv \), we take \( d = 1/\sqrt{v} \). Then the fundamental parallelogram has area one, and can be regarded as the image of the unit square \( \{(t_1, t_2) \in \mathbb{R}^2 : 0 \leq t_1 \leq 1\} \) under the linear transformation

\[
\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{v}} \begin{pmatrix} 1 & u \\ 0 & v \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix},
\]

where \( z = x + iy \) is the usual complex coordinate on \( \mathbb{C} \). A straightforward calculation gives a formula for the energy:

\[
E(f, \omega) = \frac{1}{2} \int_p \left( \left| \frac{\partial f}{\partial x} \right|^2 + \left| \frac{\partial f}{\partial y} \right|^2 \right) dx \, dy
\]

\[
= \frac{1}{2} \int_p \left( v \left| \frac{\partial f}{\partial t_1} \right|^2 + \frac{1}{v} \left| \frac{\partial f}{\partial t_2} - u \frac{\partial f}{\partial t_1} \right|^2 \right) dt_1 \, dt_2,
\]

\( P \) denoting the unit square in the coordinates \((t_1, t_2)\). Both coordinate systems \((t_1, t_2)\) and \((x, y)\) on \( \mathbb{C} \) are useful, the first respecting the product structure on the torus, the second being conformal.

Let \( P_{k,l} \) be the parallelogram described by \( 0 \leq t_1 \leq k \) and \( 0 \leq t_2 \leq l \), whenever \( k \) and \( l \) are positive integers. If \( f_0 : T^2 \to M \) is a prime minimal torus with conformal structure \( \omega \), \( f_0 \) lifts to a map \( \tilde{f}_0 : \mathbb{C} \to M \) which can then be restricted to \( P_{k,l} \). By identifying opposite edges of \( P_{k,l} \) we obtain a torus \( T^2 \) with the conformal structure \((l/k)\omega\) and \( f_0 \) induces a conformal harmonic mapping \( f_1 \) from the new torus into \( M \) which covers \( f_0 \) with multiplicity \( kl \) and has fundamental parallelogram \( P_{k,l} \).

We now carry through several constructions motivated by Bott’s treatment of multiple covers of closed geodesics [1956]. Suppose as above that \( f_0 : T^2 \to M \) is a prime minimal torus with lift \( \tilde{f}_0 : \mathbb{C} \to M \). Let \( (\tilde{f}_0^* TM)^\perp \) denote the normal bundle to the immersion \( \tilde{f}_0 \) and if \( \zeta \) and \( \omega \) are elements of \( S^1 \subset \mathbb{C} \), let

\[
\mathcal{V}(k,l),(\zeta,\omega) = \left\{ \text{ smooth sections } X \text{ of } (\tilde{f}_0^* TM)^\perp \otimes \mathbb{C} \text{ such that } X(t_1+k, t_2) = \zeta X(t_1, t_2), \ X(t_1, t_2+l) = \omega X(t_1, t_2) \right\}.
\]
Define an Hermitian inner product

\[ \langle \cdot, \cdot \rangle_{k,l} : \mathcal{V}(k,l),(z,w) \times \mathcal{V}(k,l),(z,w) \to \mathbb{C} \]

by

\[ \langle X, Y \rangle_{k,l} = \int_{P_{k,l}} \left( \frac{D X}{\partial x}, \frac{D Y}{\partial x} \right) + \left( \frac{D X}{\partial y}, \frac{D Y}{\partial y} \right) + \langle X, Y \rangle \, dx \, dy, \]

where \( D \) denotes the covariant derivative in the normal bundle defined by the Levi-Civita connection on \( M \) and the bar denotes conjugation. Note that elements of \( \mathcal{V}(k,l),(1,1) \) project to sections of \( (f_1^* TM)^\perp \otimes \mathbb{C} \), where \( f_1 \) is the torus with fundamental parallelogram \( P_{k,l} \).

**Lemma 1.** The inclusion

\[ (8) \quad \sum_{z^l = 1} \sum_{w^l = 1} \mathcal{V}(1,1),(z,w) \subset \mathcal{V}(k,l),(1,1), \]

the sum being taken over all \( k \)-th and \( l \)-th roots of unity, is an isomorphism.

**Proof:** Since the inclusion is clearly injective, it suffices to show that it is surjective. If \( X \in \mathcal{V}(k,l),(1,1) \) and \( z \) and \( w \) are \( k \)-th and \( l \)-th roots of unity, we let

\[ X_{z,w}(t_1, t_2) = \frac{1}{kl} \sum_{a=0}^{k-1} \sum_{b=0}^{l-1} z^{-a} w^{-b} X(t_1 + a, t_2 + b). \]

Then

\[ X_{z,w}(t_1 + 1, t_2) = \frac{1}{kl} \sum_{a=0}^{k-1} \sum_{b=0}^{l-1} z^{-a} w^{-b} X(t_1 + (a + 1), t_2 + b) \]

\[ = \frac{z}{kl} \sum_{a=0}^{k-1} \sum_{b=0}^{l-1} z^{-(a+1)} w^{-b} X(t_1 + (a + 1), t_2 + b) = z X_{z,w}(t_1, t_2), \]

and by a similar calculation,

\[ X_{z,w}(t_1, t_2 + 1) = w X_{z,w}(t_1, t_2), \]

so \( X_{z,w} \in \mathcal{V}(1,1),(z,w) \). Moreover, an easy calculation shows that

\[ X = \sum_{z^l = 1} \sum_{w^l = 1} X_{z,w}, \]

so the inclusion (8) is indeed an isomorphism. \( \square \)

We next define an index form

\[ I_{k,l}(\cdot, \cdot) : \mathcal{V}(k,l),(z,w) \times \mathcal{V}(k,l),(z,w) \to \mathbb{C} \]

by
(9) \[ I_{k,l}(X, \bar{Y}) = \int_{P_{k,l}} \left( (DX, D\bar{Y}) - \langle B(X) + \mathcal{K}(X), \bar{Y} \rangle \right) dA, \]

where \( D \) is the connection in the normal bundle and \( B \) and \( \mathcal{K} \) are the endomorphisms of the normal bundle appearing in the Simons formula (7). Recall that the operator \( B \) depends on the second fundamental form of \( \tilde{f}_0 \), while \( \mathcal{K} \) is linear in the components of the curvature tensor. Both \( B \) and \( \mathcal{K} \) are periodic since they are lifts of operators from \( f_0 \).

We can integrate by parts in (9), obtaining

(10) \[ I_{k,l}(X, \bar{Y}) = - \int_{P_{k,l}} \left( \langle \Delta X + B(X) + \mathcal{K}(X), \bar{Y} \rangle \right) dA = \int_{P_{k,l}} \langle L^\perp(X)\bar{Y} \rangle dA, \]

where \( \Delta \) is the Laplace operator defined by the normal connection and \( L^\perp \) is the normal Jacobi operator, \( X \) being a normal Jacobi field if and only if it satisfies the equation

(11) \[ L^\perp(X) = -\Delta X - B(X) - \mathcal{K}(X) = 0. \]

We divide the parallelogram \( P_{k,l} \) into \( kl \) parallelograms \( \{P_{a,b}\} \), where \( a \) and \( b \) are integers ranging from 1 to \( k \) and 1 to \( l \) respectively, the parallelogram \( P_{a,b} \) being defined by the conditions \( a - 1 \leq t_1 \leq a \) and \( b - 1 \leq t_2 \leq b \). Note that if \( X \in \mathcal{V}_{(1,1),(z_1,w_1)} \) and \( Y \in \mathcal{V}_{(1,1),(z_2,w_2)} \), where \( z_1 \) and \( z_2 \) are \( k \)-th roots of unity and \( w_1 \) and \( w_2 \) are \( l \)-th roots of unity, and \( X \) and \( Y \) are extended to elements of \( \mathcal{V}_{(k,l),(1,1)} \), the index form on the extensions is given by

\[ I_{k,l}(X, \bar{Y}) = \sum_{a=1}^{k} \sum_{b=1}^{l} \int_{P_{a,b}} \left( (DX, D\bar{Y}) - \langle B(X) + \mathcal{K}(X), \bar{Y} \rangle \right) dA \]

\[ = \sum_{a=1}^{k} \sum_{b=1}^{l} \int_{P_{1,1}} \left( (DX, D\bar{Y}) - \langle B(X) + \mathcal{K}(X), \bar{Y} \rangle \right) dA. \]

Thus we see that

\[ I_{k,l}(X, \bar{Y}) = \begin{cases} klI_{1,1}(X, \bar{Y}) & \text{if } z_1 = z_2 \text{ and } w_1 = w_2, \\ 0 & \text{if } z_1 \neq z_2 \text{ or } w_1 \neq w_2, \end{cases} \]

and hence the direct sum decomposition

\[ \sum_{z^a=1}^{k} \sum_{w^b=1}^{l} \mathcal{V}_{(1,1),(z,w)} \]

is orthogonal with respect to the index form \( I_{k,l} \). In particular, whenever \( z \) and \( w \) are roots of unity, the normal Jacobi operator \( L^\perp \) defined by (10) restricts to an endomorphism

\[ L^\perp: \mathcal{V}_{(1,1),(z,w)} \rightarrow \mathcal{V}_{(1,1),(z,w)}. \]
Let $N(z, w)$ denote the nullity of the index form $I_{k,l}$ restricted to $\mathcal{V}(1,1,(z,w))$,

$$N(z, w) = \dim \{ X \in \mathcal{V}(1,1,(z,w)) : I_{k,l}(X, \bar{Y}) = 0 \text{ for all } Y \in \mathcal{V}(1,1,(z,w)) \} = \dim \{ X \in \mathcal{V}(1,1,(z,w)) : L^\perp(X) = 0 \}.$$

The preceding discussion proves the following lemma, analogous to a lemma of Bott [1956] which plays a key role in his analysis of the relationship between the index and nullity of a prime smooth closed geodesic and the index and nullity of its multiple covers:

**Lemma 2.** If $z_1$ and $w_1$ are primitive $k$-th and $l$-th roots of unity,

$$\text{Nullity of } f_1 = \sum_{a=1}^{k} \sum_{b=1}^{l} N(z_1^a, w_1^b).$$

We now turn to the proof of Theorem 1 in the case of tori covering tori. Our strategy is to perturb the metric in a neighborhood of the given Morse nondegenerate minimal surface $f_0 : T^2 \to M$ in such a way that $f_0$ is preserved, but the Jacobi equations are perturbed.

We construct a variation of the Riemannian metric on $M$ of a specific form. We choose a point $p \in T^2$ such that $f^{-1}_0(f_0(p)) = \{ p \}$, and a neighborhood $U$ containing $p$ such that $f_0$ imbeds $U$ into some open set $V \subset M$. Arrange, moreover, that $V$ is the domain of local coordinates $(u_1, \ldots, u_n)$ such that $u_i(f_0(p)) = 0$ and

1. $f_0(U)$ is described by the equations $u_3 = \cdots = u_n = 0$,
2. $u_a \circ f_0 = x_a$ on $f_0(U)$, for $a = 1, 2$, where $x_1 + ix_2$ is a conformal parameter on $U$, and
3. the Riemannian metric $g$ on the ambient space takes the form $\sum g_{ij} du_i du_j$, such that when restricted to $f_0(\Sigma) \cap V$, $g_{ir} = \delta_{ir}$, for $1 \leq i \leq n$ and $3 \leq r \leq n$.

Such coordinates can be constructed using the exponential map restricted to the normal bundle of the surface $f_0(\Sigma) \cap V$ in $M$.

Following [Klingenber 1978, proof of Proposition 3.3.7], we construct a variation $\dot{g} = \sum \dot{g}_{ij} du_i du_j$ of the metric on the ambient manifold $M$ such that

$$\dot{g}_{11}(u_1, \ldots, u_n) = \dot{g}_{22}(u_1, \ldots, u_n) = \sum_{r,s=3}^{n} u_r u_s \alpha_{rs}(u_1, \ldots, u_n),$$

$$\dot{g}_{ij} = 0 \text{ if } (i, j) \neq (1, 1) \text{ or } (2, 2).$$

Here the $\alpha_{rs}$ are smooth functions which vanish outside a small tubular neighborhood of $f_0(\Sigma_0)$. A straightforward calculation shows that the resulting changes in
the Christoffel symbols

\[ \hat{\Gamma}_{kj} = \frac{1}{2} \left( \frac{\partial \hat{g}_{ik}}{\partial u_j} + \frac{\partial \hat{g}_{jk}}{\partial u_i} - \frac{\partial \hat{g}_{ij}}{\partial u_k} \right) \]

vanish except for

\[ \hat{\Gamma}_{r11} = \hat{\Gamma}_{r22} = -\sum_{s=3}^{n} u_s \alpha_{rs}, \]

\[ \hat{\Gamma}_{1r1} = \hat{\Gamma}_{11s} = \hat{\Gamma}_{2r2} = \hat{\Gamma}_{22s} = \sum_{s=3}^{n} u_s \alpha_{rs}, \text{ for } 3 \leq r \leq n. \]

We want to consider the effect of such a variation on the operator \( L^\perp \). Since the operator \( \mathcal{K} \) (which is essentially the second fundamental form of \( (f_0^* T M)^\perp \) in \( f_0^* T M \)) depends only on the Christoffel symbols along \( f_0 \) (where \( u_r = 0 \)), the variation of \( \mathcal{K} \) is zero under the metric deformation. On the other hand, the variation of the operator \( \mathcal{H} \) depends on the changes in curvature components, which are given by the formulae

\[ \hat{R}_{ijk} = \frac{\partial}{\partial u_i} (\hat{\Gamma}_{ij}^k) - \frac{\partial}{\partial u_j} (\hat{\Gamma}_{ik}^j) + \sum_m \hat{\Gamma}_{lim} \hat{\Gamma}_{mjk} - \sum_m \hat{\Gamma}_{lim} \hat{\Gamma}_{mjk} - \sum_m \hat{\Gamma}_{lim} \hat{\Gamma}_{mkj}. \]

Along \( f_0(\Sigma_0) \) all the \( \hat{\Gamma}_{ki} \) and \( \hat{\Gamma}_{ij} \) must vanish, so along \( f_0(\Sigma_0) \),

\[ \hat{R}_{1r1} = \hat{R}_{2r2} = \frac{\partial}{\partial u_r} (\hat{\Gamma}_{11s}) - \frac{\partial}{\partial u_1} (\hat{\Gamma}_{1rs}) = \alpha_{rs}. \]

Thus the normal metric variation will result in the following variation of the endomorphism \( \mathcal{H} \):

\[ \mathcal{H} \left( \sum_{r=3}^{n} h_r \frac{\partial}{\partial u_r} \right) = \sum_{r,s=3}^{n} \alpha_{rs} h_s \frac{\partial}{\partial u_r}, \]

where the \( h_r \) can be arbitrary real-valued functions.

We let \( \text{Met}(M, f_0) \) denote the space of Riemannian metrics \( g \) on \( M \) such that the one-jet \( j_1(g) \) of \( g \) agrees with the one-jet \( j_1(g_0) \) of \( g_0 \) at points of \( f_0(\Sigma_0) \). This implies that \( f_0 : T^2 \to M \) is a minimal torus for any \( g \in \text{Met}(M, f_0) \). Given any metric \( g \in \text{Met}(M, f_0) \), we have a corresponding normal Jacobi operator

\[ L_g^\perp : \Gamma((f_0^* T M)^\perp) \to \Gamma((f_0^* T M)^\perp), \]

\[ L_g^\perp = -\Delta - \mathcal{R} - \mathcal{H}(g), \]

the Laplace operator \( \Delta \) and the second fundamental form endomorphism \( \mathcal{R} \) being independent of the choice of \( g \in \text{Met}(M, f_0) \).

We now consider a given cover \( f_1 : T^2 \to M \) of \( f_0 \). We suppose, as above, that \( f_0 \) has fundamental parallelogram \( P \) while \( f_1 \) has fundamental parallelogram
By Lemma 2, the contributions to the nullity of $f_1$ come from Jacobi fields in $N(z, w)$, as $z$ and $w$ range over the $k$-th and $l$-th roots of unity. We let

$$E_k(z, w) = \text{completion of } \mathcal{V}_{(1,1), (z, w)} \text{ with respect to the } L^2_k \text{-norm,}$$

and define a map

$$F : E_k(z, w) \times \text{Met}(M, f_0)_k^2 \to E_{k-2}(z, w) \quad \text{by} \quad F(X, g) = L^\perp_g(X).$$

Note that for each choice of $g$, $X \mapsto L^\perp_g(X)$ is a linear Fredholm map of Fredholm index zero.

We let $SE_k(z, w) = \{ X \in E_k(z, w) : \| X \| = 1 \}$, the fiber of a unit sphere bundle

$$SE_k(z, w) \times \text{Met}(M, f_0)_k^2 \to \text{Met}(M, f_0)_k^2.$$

We claim that the subset

$$\mathcal{F} = \{ (X, g) \in SE_k(z, w) \times \text{Met}(M, f_0)_k^2 : L^\perp_g(X) = 0 \}$$

is a smooth submanifold. Note that any $X$ with $\| X \| = 1$ such that $L^\perp_g(X) = 0$ must be nonzero on an open dense set and hence any unit-length element of $E_{k-2}(z, w)$ not in the image of $L^\perp_g$ must be nonzero on an open dense set. Thus it follows from (12) that any element of $E_{k-2}(z, w)$ not in the image of $L^\perp_g$ is of the form $D^2 F(X, g)(\dot{g})$ for some metric variation $\dot{g} \in T_g(\text{Met}(M, f_0)_k^2)$. Thus the restriction of $F$ to the total space of the sphere bundle is a submersion, and our claim follows immediately from the implicit function theorem.

Suppose now that

$$\pi : SE_k(z, w) \times \text{Met}(M, f_0)_k^2 \to \text{Met}(M, f_0)_k^2$$

is the projection on the second factor. We claim that the restriction of $\pi$ to $\mathcal{F}$,

$$\pi : \mathcal{F} \to \text{Met}(M, f_0)_k^2,$$

is a Fredholm map of Fredholm index $-1$. To see this, we note first that

$$T_{(X, g)} \mathcal{F} = \{ (Y, \dot{g}) \in E_k(z, w) \times T_g \text{Met}(M, f_0)_k^2$$

such that $L^\perp_g(Y) + D^2 F(X, g)(\dot{g}) = 0, \langle X, Y \rangle = 0 \},$$

where $\langle \cdot, \cdot \rangle$ is the $L^2_k$ inner product. Thus $(Y, \dot{g})$ lies in the kernel of $d\pi_{(X, g)}$ if and only if $\langle X, Y \rangle = 0$ and $L^\perp_g(Y) = 0$. Since $X$ is in the kernel of $L^\perp_g$ by definition of $\mathcal{F}$, the dimension of the kernel of $d\pi_{(X, g)}$ is one less than the dimension of the kernel of the Jacobi operator $L^\perp_g$. 
We next investigate the cokernel, noting first that \( \dot{g} \mapsto D_2F(X, g)(\dot{g}) \) covers the cokernel of \( L_{\dot{g}} \). For \( Y \in E_{k-2}(z, w) \), we define a continuous linear functional

\[
T(Y) : T_g Met(M, f_0)^2_k \to \mathbb{R}
\]

by \( T(Y)(\dot{g}) = \langle (D_2F(X, g)(\dot{g}), Y)_{k-2} \rangle \), where \( \langle \cdot, \cdot \rangle_{k-2} \) denotes the \( L^2_{k-2} \) inner product. If \( \dot{g} \) is in the range of \( d\pi(X, g) \), then \( T(Y)(\dot{g}) = 0 \) whenever \( Y \) is perpendicular to the range of \( L_{\dot{g}} \). This shows that the codimension of the range of \( d\pi(X, g) \) is the dimension of the cokernel of \( L_{\dot{g}} \). Thus \( \pi \) is indeed a Fredholm map of index

\[
\dim(\text{Kernel of } L_{\dot{g}}) - 1 - \dim(\text{Cokernel of } L_{\dot{g}}) = -1.
\]

It therefore follows from the Sard–Smale theorem [Smale 1965] that for \( g \) belonging to a countable intersection of open dense subsets of \( Met(M, f_0)^2_k \), there will be no solutions \( X \) to \( L_{\dot{g}}(X) = 0 \) in \( E_k(z, w) \). There are only a countable number of tori covering a given torus, and for each covering of type \( (k, l) \) only \( kl \) choices of roots of unity \( z \) and \( w \). Therefore for a countable intersection of open dense subsets of \( Met(M, f_0)^2_k \), there will be no nonzero normal Jacobi fields for any torus covering a given nondegenerate minimal torus. This, together with the remarks at the end of the Introduction, proves Theorem 1 for tori covering tori.

4. Nonorientable surfaces

We next consider the modifications necessary to treat the case in which the prime minimal torus is replaced by a prime minimal Klein bottle \( f_0 : K^2 \to M \).

A minimal Klein bottle will be double covered by a minimal torus with a flat metric of area one that is invariant under an orientation-reversing deck transformation. Such a deck transformation is a map \( A_s : T^2 \to T^2 \) which is expressed in terms of appropriate standard coordinates \( (t_1, t_2) \) on the torus as

\[
A_s(t_1, t_2) = (t_1 + \frac{1}{2}, -t_2 - s)
\]

for \( s \in S^1 \), and consists of a translation composed with a reflection. One easily checks that this map satisfies the identity \( A_s^2 = 1 \). Recall that the Teichmüller space \( \mathcal{T} \) for the torus is the upper half-plane \( \mathbb{H} \), the point \( \omega = u + iv \in \mathbb{H} \) corresponding to the torus \( \mathbb{C}/\Lambda \), where \( \Lambda \) is generated by \( 1 \) and \( \omega \). In the case of a double cover of a Klein bottle, we arrange that the differential of \( A_s \) fixes the generator corresponding to \( 1 \) in the fundamental parallelogram, and the differential must then take \( \omega \) to \( -\omega \). Since the differential of \( A_s \) is an isometry for the flat metric, \( 1 \) and \( \omega \) must be perpendicular, and the Teichmüller space of flat Klein bottles with total area one consists of the positive imaginary numbers \( \omega = iv \) with \( v > 0 \), the fixed point set of the involution

\[
A_* : \mathbb{H} \to \mathbb{H}, \quad A_*(u + iv) = -u + iv.
\]
(See [Wolf 1967, Proposition 2.5.8] for further discussion.)

As the space of maps of Klein bottles into $M$, we can take

$$\text{Map}(K^2, M) = \{ f \in \text{Map}(T^2, M) : f \circ A_s = f \text{ for some } s \in S^1 \}.$$ 

Minimal Klein bottles can then be regarded as critical points for the restricted two-variable function

$$E : \text{Map}(K^2, M) \times \{ \omega \in \mathcal{T} : A_\ast(\omega) = \omega \} \rightarrow \mathbb{R}.$$ 

The energy is invariant under the action of $S^1 \times S^1$ on $\text{Map}(K^2, M)$ defined by

$$f(t_1, t_2) \mapsto f(t_1 + s_1, t_2 + s_2), \quad \text{for } (s_1, s_2) \in S^1 \times S^1.$$ 

In the subsequent discussion, we let $A = A_{s_0}$ for some choice of $s_0 \in S^1$, thereby breaking part of the $S^1 \times S^1$-symmetry.

Suppose now that $f : \Sigma \rightarrow M$ is an oriented double cover of a nonorientable minimal surface $f_0 : \Sigma_0 \rightarrow M$. The map $A$ induces an involution $A_\ast$ on $f^*TM$, as well as on the space of sections of $f^*TM$, and both of these actions extend complex linearly to the complexifications. Moreover, these involutions preserve both the metric and the pullback of the Levi-Civita connection. If $E = (f^*TM)^\perp \otimes \mathbb{C}$, the complexified normal bundle, the map $A_\ast$ determines a direct sum decomposition

$$\Gamma(E) = \Gamma_+(E) \oplus \Gamma_-(E),$$

where

$$\Gamma_+(E) = \{ X \in \Gamma(E) : A_\ast(X) = \bar{X} \}, \quad \Gamma_-(E) = \{ X \in \Gamma(E) : A_\ast(X) = -\bar{X} \}.$$

The sections of $\Gamma_+(E)$ can be regarded as normal deformations of the nonorientable minimal surface $f_0 : \Sigma \rightarrow M$ while the sections of $\Gamma_-(E)$ are deformation of $f$ that do not come from deformations of the underlying Klein bottle. The second variation formula (7) for area under normal variations applies immediately to sections of $E$.

A key point is that the normal Jacobi operator $L^\perp$ must commute with $A_\ast$, since $A_\ast$ preserves the normal connection and the operators $\mathcal{B}$ and $\mathcal{K}$, and hence induces maps

$$L^\perp : \Gamma_+(E) \rightarrow \Gamma_+(E) \quad \text{and} \quad L^\perp : \Gamma_-(E) \rightarrow \Gamma_-(E).$$

Since $f_0$ is nondegenerate, there are no normal Jacobi fields in $\Gamma_+(E)$, and the argument presented in Section 3 shows that for a generic choice of metric in $\text{Met}(M, f_0)^2$, there will also be no normal Jacobi fields in $\Gamma_-(E)$. The argument from Section 3 also shows that there are no Jacobi fields in any torus covering $f$ and hence in any Klein bottle or torus covering $f_0$. This finishes the proof of the theorem for Klein bottles.

Finally, it remains only to establish the second statement in the Theorem, and this is a relatively straightforward modification of the previous argument.
Suppose that \( f : \Sigma \to M \) is an oriented double cover of a nonorientable minimal surface \( f_0 : \Sigma_0 \to M \) of arbitrary genus, with \( A \) being the sheet interchange map, so \( f \circ A = f \). As in the special case of the Klein bottle, the map \( A \) induces an involution \( A^* \) on the complexified normal bundle \( E = (f^* TM) \perp \mathbb{C} \) of \( f \), as well as on the space of sections of \( E \), and we have a direct sum decomposition (14). Just as before, there are no Jacobi fields in \( \Gamma(E_+) \) since \( f_0 \) is assumed to be Morse nondegenerate, and the argument presented in Section 3 shows that for generic choice of metric in \( \text{Met}(M, f_0)^2 \), there will also be no Jacobi fields in \( \Gamma(E_-) \). This finishes the proof of Theorem 1.

5. Bumpy metrics for harmonic maps

If \( f \) is an \( \omega \)-harmonic map, the \textit{Hopf differential} is the holomorphic quadratic differential
\[
\Omega_f = \left( \frac{\partial f}{\partial z}, \frac{\partial f}{\partial \bar{z}} \right) dz^2,
\]
and it vanishes precisely when \( f \) is conformal (that is, it satisfies (6)). Note that \( \Omega_f \) automatically vanishes at branch points.

In the case where \( \Sigma \) is the torus \( T^2 \), \( \Omega_f = adz^2 \), where \( a \) is a complex constant. If \( f \) is not conformal, \( a \neq 0 \), and \( f \) cannot have any branch points. On the other hand, if \( a = 0 \), it follows from the Main Theorem of [Moore 2006] that \( f \), now a parametrized minimal surface, has no branch points for generic choice of Riemannian metric on \( M \), when \( M \) has dimension at least four.

However, in contrast with minimal tori, nonconformal \( \omega \)-harmonic tori \( f : T^2 \to M \) can have points at which the rank of \( df \) is one. This can happen in one of two ways: \( f \) can be a torus parametrization of a smooth closed geodesic, or \( f \) can have fold points.

Example. To see how the second case arises, we consider a degree zero harmonic map \( f : T^2 \to S^2 \), where \( S^2 \) is given the standard Riemannian metric of constant curvature one, which has “fold points” along two circles parallel to the equator. To construct \( f \), we first note that the metric on \( S^2 \subset \mathbb{R}^3 \) with equation \( x^2 + y^2 + z^2 = 1 \) is expressed in spherical coordinates \((\phi, \theta)\), where \( z = \cos \phi \) and \( \theta \) is the standard angular coordinate in the \((x, y)\)-plane, as
\[
ds^2 = (\cos^2 \phi) d\theta^2 + d\phi^2 = \text{sech}^2 u (d\theta^2 + du^2),
\]
where \( u \) and \( \phi \) are related by the equation \( \tanh(u/2) = \tan(\phi/2) \). In terms of the standard coordinates \((t_1, t_2)\) on \( T^2 \), we can define a mapping \( f : T^2 \to S^2 \) by
\[
\theta(t_1, t_2) = t_2, \quad \phi(t_1, t_2) = \phi(t_1),
\]
where \( \phi \) is a (nonconstant speed) parametrization of the geodesic \( \theta = \text{constant} \).
The circle \( \phi = \text{constant} \) has curvature \( \kappa = 1/\cos \phi \) and normal curvature \( \kappa_n = 1 \). The equation \( \kappa^2 + \kappa_n^2 = \kappa^2 \), where \( \kappa \) is the geodesic curvature, implies that \( \kappa_n = \pm \tan \phi \). Moreover, the curve is traversed with constant speed \( \cos \phi \). Hence

\[
0 = \frac{D}{dt_1} \left( \frac{\partial f}{\partial t_1} \right) + \frac{D}{dt_2} \left( \frac{\partial f}{\partial t_2} \right) = \frac{d^2 \phi}{dt^2} + (\tan \phi)(\cos^2 \phi) = \frac{d^2 \phi}{dt^2} + \frac{1}{2} \sin 2\phi.
\]

Thus the differential equation we must solve to obtain a harmonic map (the pendulum equation except for constant factors) is equivalent to the first order system

\[
d\phi/dt = \psi, \quad d\psi/dt = \frac{1}{2} \sin 2\phi.
\]

Eliminating \( dt \) yields

\[
\frac{d\psi}{d\phi} = \frac{\frac{1}{2} \sin 2\phi}{\psi}, \quad \text{which integrates to} \quad \frac{1}{2} \psi^2 + \frac{1}{2} \cos 2\phi = (\text{constant}).
\]

For any constant strictly less than one-half we get solutions to the differential equations which yield harmonic maps for appropriate conformal structures on \( T^2 \). As the constant approaches one-half, the conformal structure approaches the boundary of Teichmüller space for the torus.

The antipodal map \( A : S^2 \to S^2 \) induces an orientation reversing map \( A : T^2 \to T^2 \) such that \( f \circ A = A \circ f \). We can take the quotient in both domain and range, obtaining thereby a harmonic map from a Klein bottle into the real projective plane \( \mathbb{R}P^2 \), which has as its image a Möbius band.

Thus it is possible to construct four types of nonconstant harmonic tori without branch points which are not branched covers of minimal spheres, do not degenerate to geodesics, and do not cover harmonic tori of lower energy: immersed tori, double covers of immersed Klein bottles, harmonic cylindrical bands and double covers of harmonic Möbius bands.

The isotropy group \( \Gamma \) of the \( O(2) \times O(2) \)-action on \( \text{Map}(T^2, M) \) is trivial for immersed tori and \( Z_2 \) for harmonic cylindrical bands, the generator being a reflection in one of the two \( O(2) \)-factors. The isotropy group is \( Z_2 \) for Klein bottles and \( \mathbb{Z}_2 \oplus \mathbb{Z}_2 \) for harmonic Möbius bands.

We say that a harmonic surface \( f : \Sigma \to M \) covers a harmonic surface \( f_0 : \Sigma_0 \to M \) if there is a conformal map \( g : \Sigma \to \Sigma_0 \) such that \( f = f_0 \circ g \).

**Definition.** An \( \omega \)-harmonic map \( f : \Sigma \to M \) is prime if it is nonconstant and is not a cover (with possible branch and fold points) of an \( \omega' \)-harmonic map \( f_0 : \Sigma_0 \to M \) of lower energy, the surfaces \( \Sigma \) and \( \Sigma_0 \) being allowed to be nonorientable.

**Theorem 2.** Suppose \( M \) is a compact connected smooth manifold of dimension at least four with a generic choice of Riemannian metric. Then the nonconstant prime \( \omega \)-harmonic tori and Klein bottles are free of branch points. Moreover, either
(1) they are parametrizations of smooth closed geodesics and lie on one-di-

dimensional nondegenerate critical submanifolds, or

(2) they lie on two dimensional critical submanifolds, each an orbit for the action

of $O(2) \times O(2)$ on $\text{Map}(T^2, M)$.

The same holds for all $\omega$-harmonic tori and Klein bottles which are unbranch-

ed covers of $\omega'$-harmonic tori and Klein bottles.

When the dimension of $M$ is three, a version of Theorem 2 holds for noncon-

formal harmonic maps.

In Theorem 2, we allow the possibility that the minimal surfaces may have fold

points, points at which the fiber of the line bundle $L$ coincides with its conjugate.

It would be interesting to determine whether fold points exist in manifolds of di-

mension at least four with generic metrics.

In analogy with Theorem 1, the motivation behind Theorem 2 is that it is needed

to provide part of the foundation for a partial parametrized Morse theory for $\omega$-

harmonic maps, the conformal structure $\omega$ being the parameter.

6. Proof of Theorem 2

The proof of Theorem 2 is similar to that of Theorem 1, except that we use second

variation of $E_\omega$ instead of second variation of $A$. Once one proves Theorem 2 for

prime harmonic maps, the extension to unbranched covers is proven in exactly the

same way as in Sections 3 and 4.

If $f$ is conformal, it follows from the Main Theorem of [Moore 2006] that $f$ is

free of branch points. If $M$ is nonconformal, it cannot have branch points, as we

already mentioned. Proposition 3.1 of the same paper shows that the theorem holds

for those prime harmonic maps $f$ which are free of branch points and somewhere

injective, that is satisfy the condition that $f^{-1}(f(p)) = p$ for some $p \in T^2$. Thus

in the prime orientable case, we need only analyze the prime $\omega$-harmonic maps

$f : T^2 \to M$ which fail to be somewhere injective.

Following the proof of Theorem 3 in [Sampson 1978], we note that it follows

from Aronsjazn’s unique continuation theorem for harmonic maps that if $df$ has

rank zero on a nonempty open set, the harmonic map $f$ must be constant. If $df$ has rank one on a nonempty open set $U \subset \Sigma$, every point of $U$ has an open neighborhood which is mapped by $f$ onto a smooth arc $C$ in $M$. We can suppose that coordinates $(u, \theta)$ have been constructed on $U$ so that $\partial f / \partial \theta = 0$, and thus $f : U \to M$ reduces to a function of one variable, $f(u, \theta) = f_0(u)$, parametrizing a curve $C$. In this case the harmonic map must be a parametrization of a geodesic.

Finally, we need to analyze the case in which $df$ has rank two on an open set. This includes harmonic maps with fold points at which the rank drops to one.
To analyze such maps, we let

\[ F = \{ p \in \Sigma : L(p) = \overline{L(p)} \} , \]

points at which the rank of \( df \) is one, and carry through the theory of branched immersions described in [Gulliver et al. 1973], allowing now, however, for the possibility of folding of \( f \) along \( F \).

We can carry out the analysis for an arbitrary Riemann surface, not just a torus or Klein bottle. We begin by defining an equivalence relation \( \sim \) on points of \( \Sigma - F \) by setting \( p \sim q \) if there is an open neighborhoods \( U_p \) and \( U_q \) of \( p \) and \( q \) respectively, and a conformal or anticonformal diffeomorphism \( \psi : U_p \rightarrow U_q \) such that \( f \circ \psi = f \). Using the argument in [Gulliver et al. 1973], which is based upon Aronsjazn’s unique continuation theorem, one shows that \( \sim \) is indeed an equivalence relation and that the quotient space \( \Sigma_0 = \Sigma - F / \sim \) is a smooth manifold except at branch points if they exist. Moreover, the conformal structure \( \omega \) on \( \Sigma \) projects to a conformal structure \( \omega_0 \) on \( \Sigma_0 \). (In the case where \( \Sigma \) is a torus and \( f \) is not conformal, \( f \) has no branch points, of course.) We can define \( f_0 : \Sigma_0 \rightarrow M \) by \( f_0([p]) = f(p) \), where \([p]\) denotes the equivalence class of \( p \), so that if \( \pi : \Sigma \rightarrow \Sigma_0 \) is the quotient map, \( f_0 \circ \pi = f \). We note that any point equivalent to a branch point is itself a branch point. The restriction of \( f_0 \) to \( \Sigma_0 \) minus the equivalence classes of the branch points is a harmonic map of finite energy. It therefore follows from the removable singularity theorem of Sacks and Uhlenbeck [1981, Theorem 3.6] that the restriction of \( f_0 \) can be extended to the equivalence classes of the branch points so as to be a harmonic map.

If \( f \) has fold points, \( \Sigma_0 \) will consist of several connected components (at least two). Each such component will be diffeomorphic to a component of \( \Sigma - F \). Moreover, two components of \( \Sigma - F \) which have the same components of \( F \) in their closure must be diffeomorphic. In the case of the torus this implies that no component of the set \( F \) of fold points can be a null homotopic circle, because this would imply that a disk is diffeomorphic to something which is not a disk. In the case where \( \Sigma \) is a torus, all of the components of \( F \) must be smooth closed circles and must all lie in the same homology class.

Note that the complex dilatation \( K_f(p) \) as described in [Bers 1960] or [Imayoshi and Taniguchi 1992] must go to infinity at points of \( F \). Moreover, \( K_{f_0}([p]) = K_f(p) \). It follows that all components of \( \Sigma - F \) are diffeomorphic to each other. Moreover, if there were more than two components, then \( f \) would not be prime, and hence there are exactly two sheets to the covering \( \pi : \Sigma - F \rightarrow \Sigma_0 \).

We can define a map \( B : \Sigma \rightarrow \Sigma \) which fixes \( F \) and interchanges the two sheets of the cover; thus \( f \circ B = f \). The map \( B \) is an orientation-reversing isometry of \( T^2 \) with its flat metric, and it induces a map (also denoted by \( B \)) from \( L \) to \( \overline{L} \) such that \( B \circ \pi = \pi \circ B \), where \( \pi : L \rightarrow \Sigma \) is the projection.
The isometry \( B : T^2 \to T^2 \) induces a conjugate linear involution
\[
B_* : E \to E, \quad \text{where} \quad E = f^*TM \otimes \mathbb{C},
\]
and we divide the space \( \Gamma(E) \) of sections of \( E \) into a direct sum \( \Gamma(E) = \Gamma_+(E) + \Gamma_-(E) \), where
\[
\Gamma_+(E) = \{ X \in \Gamma(E) : B_*(X) = \bar{X} \}, \quad \Gamma_-(E) = \{ X \in \Gamma(E) : B_*(X) = -\bar{X} \}.
\]
The sections of \( \Gamma_+(E) \) can be seen as deformations of the harmonic cylindrical band, while the sections of \( \Gamma_-(E) \) are deformations of \( f \) that separate the two sheets of \( \Sigma_0 \).

We can now apply the argument of Section 4 to \( f \). Once again, we see that variations in the metric on \( M \) eliminate all of the potential Jacobi fields for the function \( E_\omega \). Thus an application of the Sard–Smale theorem [Smale 1965] shows that for generic choice of metric there are no Jacobi fields tangent to the surface except for the those generated by the action of the symmetry group \( S^1 \times S^1 \).

The argument for the case of a prime harmonic Klein bottle \( f : K^2 \to M \) is similar. In this case, \( f \) is double covered by a harmonic torus \( \tilde{f} : T^2 \to M \) with covering transformation \( A \). If \( f \) is not somewhere injective, we construct \( \Sigma_0 \) as above, but now \( \Sigma_0 \) has only one component and the isometry \( B : K^2 \to K^2 \) fixing \( F \) lifts to an isometry \( B : T^2 \to T^2 \) on the double cover which commutes with \( A \).

We can now divide the space \( \Gamma(E) \) into a direct sum
\[
\Gamma(E) = \Gamma_{++}(E) + \Gamma_{+-}(E) \oplus \Gamma_{-+}(E) + \Gamma_{--}(E),
\]
where, for example,
\[
\Gamma_{+-}(E) = \{ X \in \Gamma(E) : A_*(X) = \bar{X}, B_*(X) = -\bar{X} \}.
\]
the other summands being defined similarly. Once again, we apply the previous argument and the Sard–Smale theorem to finish the proof of Theorem 2.

Acknowledgment

The author takes pleasure in thanking the referee for helpful suggestions leading to an improvement in the article.

References


Received August 17, 2005. Revised February 13, 2007.

JOHN DOUGLAS MOORE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
SANTA BARBARA, CA 93106
UNITED STATES
moore@math.ucsb.edu
http://www.math.ucsb.edu/~moore
UNRAMIFIED 3-EXTENSIONS
OVER CYCLIC CUBIC FIELDS

AKITO NOMURA

We study the existence of unramified 3-extensions over cyclic cubic fields. As an application, we study the class number relation between certain cubic fields.

1. Introduction

Let $F$ be a number field and $\Gamma$ a finite group. We are interested in the problem whether there exists an unramified Galois extension $M/F$ with Galois group isomorphic to $\Gamma$. In case when $\Gamma$ is an abelian group, by class field theory, this problem is closely related to the structure of the ideal class group of $F$. Thus this problem is interesting in the sight of a generalization of class field theory.

In this article we consider the following problems.

Problem $P(F, \Gamma)$: For a given Galois extension $F/\mathbb{Q}$ and a finite group $\Gamma$, does there exists a Galois extension $M/F/\mathbb{Q}$ satisfying the conditions:

1. $\text{Gal}(M/F)$ is isomorphic to $\Gamma$;
2. $M/F$ is unramified?

By definition, “a Galois extension $M/F/\mathbb{Q}$” means that $M/\mathbb{Q}$, $F/\mathbb{Q}$ are Galois extensions, with $F$ an intermediate field of $M/\mathbb{Q}$.

Problem $P(F, \Gamma, E)$: For a given Galois extension $F/\mathbb{Q}$ and finite groups $\Gamma$ and $E$, does there exists a Galois extension $M/F/\mathbb{Q}$ satisfying the conditions:

1. $\text{Gal}(M/F)$ is isomorphic to $\Gamma$;
2. $\text{Gal}(M/\mathbb{Q})$ is isomorphic to $E$;
3. $M/F$ is unramified?

If a Galois extension $M/F/\mathbb{Q}$ satisfies the conditions in $P(F, \Gamma)$, we call the field $M$ a solution of $P(F, \Gamma)$, and likewise for $P(F, \Gamma, E)$.

MSC2000: 12F12, 11R29.

Keywords: cyclic field, cubic field, unramified extension, 3-extension, class number.

This research was partially supported by the Grants-in-Aid for Scientific Research (C), the Japan Society for the Promotion of Science.
In [Nomura 1991; 1993; 2002], we studied these problems in the case where \( l \) and \( p \) are distinct primes, \( F \) is a cyclic field of degree \( l \), and \( \Gamma \) is a \( p \)-group. Lemmermeyer [1997] conjectured that for any 2-group \( \Gamma \) there exists a quadratic field \( F \) such that the answer to the problem \( P(F, \Gamma) \) is affirmative, but this has been disproved by Boston and Leedham-Green [1999].

Here we shall study the problems above for cyclic cubic fields and certain 3-groups. As an application of our main result, we study the class number relations of some cubic fields and the class number of the Hilbert 3-class field of certain cubic fields. We also provide an alternative proof for a part of the result in [Naito 1987] and a slight generalization. We use GAP Version 4.4 for calculations of 3-groups.

2. Preliminary from embedding problems

In this section, we quote some results about embedding problems. General studies on embedding problems can be found in [Hochsmann 1968; Neukirch 1973].

Let \( \mathfrak{G} \) be the absolute Galois group of a number field \( k \), and \( L/k \) a finite Galois extension with Galois group \( G \). For a central extension

\[
\varepsilon : 1 \rightarrow A \rightarrow E \xrightarrow{j} G \rightarrow 1,
\]

the embedding problem \((L/k, \varepsilon)\) is defined by the diagram

\[
\begin{array}{cccccc}
\mathfrak{G} & \downarrow \varphi & & & & \\
\varepsilon : 1 & \rightarrow & A & \rightarrow & E & \rightarrow \ G & \rightarrow & 1,
\end{array}
\]

where \( \varphi \) is the canonical surjection. A continuous homomorphism \( \psi \) of \( \mathfrak{G} \) to \( E \) is called a solution of \((L/k, \varepsilon)\) if it satisfies the condition \( j \circ \psi = \varphi \). When \((L/k, \varepsilon)\) has a solution, we call \((L/k, \varepsilon)\) is solvable. A solution \( \psi \) is called a proper solution if it is surjective. A field \( M \) is also called a solution (resp. proper solution) of \((L/k, \varepsilon)\) if \( M \) is corresponding to the kernel of any solution (resp. proper solution).

For each prime \( q \) of \( k \), we write \( k_q \) for the \( q \)-completion of \( k \), and \( L_q \) for the completion of \( L \) relative to an extension of \( q \) to \( L \). The local problem \((L_q/k_q, \varepsilon_q)\) of \((L/k, \varepsilon)\) is defined by the diagram

\[
\begin{array}{cccccc}
\mathfrak{G}_q & & \downarrow \varphi_{|\varepsilon_q} & & & & \\
\varepsilon_q : 1 & \rightarrow & A & \rightarrow & E_q & \xrightarrow{j|_{E_q}} & G_q & \rightarrow & 1,
\end{array}
\]

where \( G_q \) is the Galois group of \( L_q/k_q \), which is isomorphic to the decomposition group of \( q \) in \( L/k \), \( \mathfrak{G}_q \) is the absolute Galois group of \( k_q \), and \( E_q \) is the inverse of
$G_q$ by $j$. In the same manner as the case of $(L/k, \varepsilon)$, solution and proper solution are defined for $(L_q/k_q, \varepsilon_q)$.

We need some lemmas, which are essential in the theory of embedding problems. Let $p$ be an odd prime and $L/k$ a $p$-extension. Let $\varepsilon : 1 \to \mathbb{Z}/p\mathbb{Z} \to E \to \text{Gal}(L/\mathbb{Q}) \to 1$ be a central extension.

We denote by $\text{Ram}(L/k)$ the set of all primes of $k$ which are ramified in $L/k$.

**Lemma 2.1** [Neukirch 1973]. $(L/k, \varepsilon)$ is solvable if and only if $(L_q/k_q, \varepsilon_q)$ are solvable for all primes $q$ of $\text{Ram}(L/k)$.

**Lemma 2.2** [Hoechsmann 1968]. If $\varepsilon$ is a nonsplit extension, every solution of $(L/k, \varepsilon)$ is a proper solution.

**Lemma 2.3** [Neukirch 1973]. Assume that $(L/k, \varepsilon)$ is solvable. Let $S$ be a finite set of primes of $k$ and $M(q)$ a solution of $(L_q/k_q, \varepsilon_q)$ for $q$ of $S$. Then there exists a solution $M$ of $(L/k, \varepsilon)$ such that the completion of $M$ by $q$ is equal to $M(q)$ for each $q$ of $S$.

### 3. Embedding problems with ramification conditions

Let $p$ be an odd prime. In this section, let $k$ be either the rational number field or an imaginary quadratic field with the class number prime to $p$ ($p \neq 3$, when $k = \mathbb{Q}(\sqrt{-3})$).

We now state a key lemma of this article. The idea of the proof is similar to [Nomura 1991], and we sketch it for the reader’s convenience.

**Lemma 3.1.** Let $L/k$ be a $p$-extension and $\varepsilon : 1 \to \mathbb{Z}/p\mathbb{Z} \to E \to \text{Gal}(L/k) \to 1$ a nonsplit central extension. Assume that the induced extension $\varepsilon_q$ is split for any prime $q$ of $\text{Ram}(L/k)$. Then $(L/k, \varepsilon)$ has a proper solution $M$ such that $M/L$ is unramified.

**Proof.** For any prime $q$ of $\text{Ram}(L/k)$, the local problem $(L_q/k_q, \varepsilon_q)$ is solvable because $\varepsilon_q$ is split. By Lemma 2.1, $(L/k, \varepsilon)$ is solvable.

Next we shall prove that for each prime $p$ of $k$ above $p$ the local problem $(L_p/k_p, \varepsilon_p)$ has a solution $M(p)/L_p/k_p$ such that $M(p)/L_p$ is unramified. If $\varepsilon_p$ is split, then $L_p$ is itself a solution. Assume that $\varepsilon_p$ is not split. Then $p$ is unramified in $L/k$, and $\text{Gal}(L_p/k_p)$ is cyclic $p$-group. Hence $E_p$ is also cyclic $p$-group. Since the Galois group of the maximal unramified $p$-extension of $k_p$ is isomorphic to the ring of $p$-adic integers, the problem $(L_p/k_p, \varepsilon_p)$ has an unramified solution.

By virtue of Lemma 2.2 and 2.3, $(L/k, \varepsilon)$ has a proper solution $M_1/L/k$ such that any prime $\overline{p}$ of $L$ above $p$ is unramified in $M_1/L$. If $M_1/L$ is unramified, $M_1/L/k$ is a required solution of $(L/k, \varepsilon)$. Assume that $M_1/L$ is not unramified. Let $\widehat{q}$ be a prime of $M_1$ which is ramified in $M_1/L$ and $\widehat{q}$ (resp. $q$) the restriction
to $L$ (resp. $k$). Then $N_{M_1/q\bar{q}} \equiv 1 \mod p$. Since $M_1/k$ is a $p$-extension, $N_{k/q\bar{q}} \equiv 1 \mod p$. By [Shafarevich 1964, Theorem 1], there exists an extension $T/k$ such that $q$ is ramified in $T/k$ and that other primes are unramified. Let $\bar{q}$ be an extension of $q$ to $M_1T$ and $M_2$ the inertia field of $\bar{q}$ in $M_1T/k$. By the assumption of $\epsilon_q$, $q$ is unramified in $L/k$ because the inertia group of $\bar{q}$ in $M_1/k$ is cyclic. Then $M_2$ is a proper solution of $(L/k, \epsilon)$ such that $\text{Ram}(M_2/L) \subseteq \text{Ram}(M_1/L)$. By repeating this process, we can get a required solution. □

4. Lemmas on $p$-extensions

In this section we shall prepare some lemmas and notations.

For each odd prime $p$, denote by $E(p^3)$ the group of order $p^3$ defined by

$$\langle x, y, z \mid x^p = y^p = z^p = 1, x^{-1}yx = yz, xz = zx, yz = zy \rangle.$$ 

The next two lemmas are essential in this article. Lemma 4.2 is a special case of the Chebotarev monodromy theorem; for the proof see [Cohn 1978, Theorem 16.30].

**Lemma 4.1.** Let $k$ be a number field and $M/L/k$ a Galois extension such that

1. $\text{Gal}(M/k) \cong E(p^3)$,
2. $\text{Gal}(L/k) \cong \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$,
3. $M/L$ is unramified.

Then $L/k$ is locally cyclic, that is to say, any prime ramified in $L/k$ is also decomposed in $L/k$.

**Proof.** Assume that there exists a prime $q$ of $k$ such that $\text{Gal}(L_q/k_q) \cong \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. Let $\tilde{q}$ and $\bar{q}$ be primes of $M$ and $L$, respectively, above $q$. We must consider two cases. First assume that $q$ is totally ramified in $L/k$. We remark that this case occur only when $q$ is above $p$. Since $M/L$ is unramified, the order of the inertia group of $\bar{q}$ in $M/k$ is $p^2$. Then the inertia group is normal subgroup of $\text{Gal}(M/k)$, so the inertia field is a cyclic extension over $k$ of degree $p$. Hence it is contained in $L$. This is a contradiction. Next assume that $q$ is inert and ramified in $L/k$. Since $E(p^3)$ has no cyclic subgroup of order $p^2$, $\bar{q}$ is decomposed in $M/L$. Then the order of the decomposition group of $\tilde{q}$ in $M/k$ is $p^2$. Thus the decomposition group is normal subgroup of $\text{Gal}(M/k)$. Hence the decomposition field is contained in $L$. This is a contradiction. □

**Lemma 4.2.** Let $p$ be a prime and $k$ a number field such that the class number is prime to $p$. Let $F/k$ be a cyclic extension of degree $p$. If $L/F/k$ is a $p$-extension such that $L/F$ is unramified, then $\text{Gal}(L/k)$ is generated by elements of degree $p$. 
Notation. In the rest of this article, we write $\Gamma(i, j)$ for the group whose library number in GAP is $(i, j)$, where $i$ is equal to the order of its group. With the commutator notation $[\alpha, \beta] = \alpha^{-1} \beta^{-1} \alpha \beta$ and the ordinary generator-relator notation, we have
\[
\begin{align*}
\Gamma(3, 2) & = \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \\
\Gamma(3, 3) & = \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \\
\Gamma(3, 4) & = \langle x, y, z \mid x^3, y^3, z^3, [x, y, z], [y, z] \rangle = E(3^3), \\
\Gamma(3, 5) & = \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \\
\Gamma(3, 6) & = \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z}, \\
\Gamma(3, 7) & = \langle x, y, z \mid y[x, x], x^9, y^3, z^3, [x, y], [z, y] \rangle, \\
\Gamma(3, 8) & = \langle x, y \mid x^9, y^3, x^3[y, x] \rangle, \\
\Gamma(3, 9) & = \langle x, y, z \mid y[x, x], x^9, y^3, z^3, x^3[y, z], [x, y] \rangle, \\
\Gamma(3, 10) & = \langle x, y, z \mid y[x, x], x^9, y^3, x^3[y, x], z^3x^3, [y, z] \rangle, \\
\Gamma(3, 12) & = \Gamma(3, 3) \times \mathbb{Z}/3\mathbb{Z}, \\
\Gamma(4, 2) & = \langle x, y, z, u, v \mid z[x, y], x^3u^{-1}, y^3v^{-1}, z^3, u^3, v^3, [x, z], [y, z], [y, u], [x, v] \rangle, \\
\Gamma(4, 3) & = \langle x, y, z, u, v \mid z[x, y], u[x, z], v[y, z], x^3, y^3, z^3, u^3, v^3, [y, u], [z, u], [x, u], [y, v], [z, v], [u, v], [x, v] \rangle, \\
\Gamma(4, 7) & = \langle x, y, z \mid y[x, x], u[x, z], v[y, z], x^3u^{-1}, y^3v, z^3, u^3, v^3, [y, z], [y, u], [z, u], [x, v], [z, v] \rangle, \\
\Gamma(4, 9) & = \langle x, y, z \mid y[x, x], u[x, z], v[y, z], x^3u^{-1}, y^3v^3, z^3, u^3, v^3, [y, z], [y, u], [z, u], [x, v], [y, v] \rangle, \\
\Gamma(4, 10) & = \langle x, y, z \mid y[x, x], u[x, z], v[y, z], x^3u^{-1}, y^3v, z^3u, v^3, [z, u], [x, v], [y, v] \rangle, \\
\Gamma(5, 15) & = \langle x, y, z, u, v \mid z[x, y], u[x, z], v[y, z], x^3u^{-1}, y^3v^3, z^3u^3, (xy)^3, [x, z], [y, u], [z, u], [x, v], [y, v] \rangle, \\
\Gamma(5, 26) & = \langle x, y, z, u, v \mid u[x, z], v[y, z], z[x, y], x^3u, v^3, z^3u^3v, [z, u], [x, v], [y, v] \rangle, \\
\Gamma(5, 28) & = \langle x, y, z, u, v \mid u[x, z], v[y, z], z[x, y], x^3u^3, v^3, z^3u^3v, [z, u], [x, v], [y, v] \rangle, \\
\Gamma(5, 33) & = \langle x, y, z, u, v \mid u[x, y], v[x, u], y^3v, x^3u^3, z^3u^3v, [x, z], [y, z], [y, u], [z, u], [x, v], [u, v] \rangle, \\
\Gamma(6, 40) & = \langle x, y, z, u, v, w \mid v[y, z], u[x, z], v[x, w], z[x, y], z^3w, x^3, y^3, u^3, w^3, [z, v], [u, v], [z, u], [y, w], [u, w], [v, w], [x, w] \rangle,
\end{align*}
\]
Using GAP, we locate all nonabelian 3-groups \( \Gamma \) satisfying three conditions:

(G1) \( \Gamma \) is generated by elements of order 3.
(G2) The 3-rank of \( \Gamma \) is equal to 2.
(G3) The order of \( \Gamma \) is between \( 3^2 \) and \( 3^5 \).

We list in Table 1 their maximal subgroups. By condition (G2), there are always four of them.

\[
\begin{array}{|c|c|}
\hline
\Gamma & \text{maximal subgroups of } \Gamma \\
\hline
\Gamma(3^3, 3) & \Gamma(3^2, 2) \times 4 \\
\Gamma(3^4, 7) & \Gamma(3^3, 3), \ \Gamma(3^3, 4) \times 2, \ \Gamma(3^3, 5) \\
\Gamma(3^4, 9) & \Gamma(3^3, 2), \ \Gamma(3^3, 3) \times 3 \\
\Gamma(3^5, 3) & \Gamma(3^4, 3) \times 2, \ \Gamma(3^4, 12) \times 2 \\
\Gamma(3^5, 26) & \Gamma(3^4, 2), \ \Gamma(3^4, 9) \times 3 \\
\Gamma(3^5, 28) & \Gamma(3^4, 4), \ \Gamma(3^4, 9) \times 2, \ \Gamma(3^4, 10) \\
\hline
\end{array}
\]

Table 1. 3-groups satisfying conditions (G1), (G2), and (G3). The notation \( \Gamma(i, j) \times r \), for \( r > 1 \), means that there exist \( r \) maximal subgroups isomorphic to \( \Gamma(i, j) \).

Let \( L/F/\mathbb{Q} \) be a Galois extension such that \( F/\mathbb{Q} \) is a cyclic cubic extension and \( L/F \) is an unramified 3-extension. Then by Lemma 4.2, \( \text{Gal}(L/\mathbb{Q}) \) must satisfy condition (G1).

**Remark 4.3.** Let \( x, y, z \) be generators of \( \Gamma(3^4, 9) \) as in the presentation of the previous page. The maximal subgroups of \( \Gamma(3^4, 9) \) are \( \langle x, y \rangle, \ \langle y, z \rangle, \ \langle xz, y \rangle, \) and \( \langle x^2z, y \rangle \), where the first is isomorphic to \( \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \) and the others are isomorphic to \( \Gamma(3^3, 3) \). If we replace \( xz \) (or \( x^2z \)) by \( z \), then \( x, y, z \) satisfy the same relations as in the original presentation.

5. Unramified 3-extensions over cyclic cubic fields

Let \( F/\mathbb{Q} \) be a cyclic cubic extension. For some finite 3-groups \( \Gamma \) and \( E \), we shall consider the problems \( P(F, \Gamma) \) and \( P(F, \Gamma, E) \) defined in the Introduction.

First we define some conditions concerning the Galois extension \( L_0/F/\mathbb{Q} \):

(C1) \( \text{Gal}(L_0/\mathbb{Q}) \) is isomorphic to \( \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \).
(C2) \( L_0/\mathbb{Q} \) is locally cyclic.
(C3) \( L_0/F \) is an unramified cubic extension.
There exists a cubic subfield $F'$ of $L_0$ such that $F' \neq F$ and that $L_0/F'$ is unramified.

**Remark 5.1.** Under (C1), condition (C2) is equivalent to that any prime of $\mathbb{Q}$ ramified in $L_0/\mathbb{Q}$ is decomposed in $L_0/\mathbb{Q}$.

**Remark 5.2.** Assume that $L_0/F/\mathbb{Q}$ satisfies conditions (C1), (C2) and (C3). If only two primes of $\mathbb{Q}$ are ramified in $F/\mathbb{Q}$, then condition (C4) is always satisfied.

**Proposition 5.3.** Assume that the Galois extension $L_0/F/\mathbb{Q}$ satisfies the conditions (C1) and (C3). There is equivalence between

(a) $L_0/F/\mathbb{Q}$ satisfies condition (C2);

(b) $P(F, \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^3, 3))$ has a solution $L_1$ such that $L_1 \supset L_0$.

**Proof.** The implication (b) $\Rightarrow$ (a) is clear by Lemma 4.1. We shall prove (a) $\Rightarrow$ (b). There exists a nonsplit central extension

$$\epsilon : 1 \to \mathbb{Z}/3\mathbb{Z} \to \Gamma(3^3, 3) \to \text{Gal}(L/\mathbb{Q}) \to 1.$$ 


The explicit construction of $\epsilon$ is as follows. Let $F'$ be an any cubic subfield of $L_0$ such that $F' \neq F$, and put $\text{Gal}(L_0/F) = \langle a \rangle$, $\text{Gal}(L_0/F') = \langle b \rangle$. Let $\Gamma(3^3, 3) = \langle x, y, z | x^3, y^3, z^3, z[y, x], [x, z], [y, z] \rangle$. Then $j$ is defined by $x \mapsto a, y \mapsto b.$

Since the exponent of the group $\Gamma(3^3, 3)$ is equal to 3, the induced extension $\epsilon_q$ is split for any prime $q$. By applying Lemma 3.1 to the embedding problem $(L_0/\mathbb{Q}, \epsilon)$, we can find a Galois extension $L_1/L_0/\mathbb{Q}$ such that $\text{Gal}(L_1/\mathbb{Q})$ is isomorphic to $\Gamma(3^3, 3)$ and that $L_1/L_0$ is unramified. Since $L_0/F$ is unramified, $L_1/F$ is also unramified. Further $\text{Gal}(L_1/F) = j^{-1}(\langle a \rangle) = \langle x, z \rangle \cong \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$. 

**Corollary 5.4.** Let $q$ and $l$ be prime numbers such that $q \equiv l \equiv 1 \text{ mod } 3$, $q^{(l-1)/3} \equiv 1 \text{ mod } l$, and $l^{(q-1)/3} \equiv 1 \text{ mod } q$. Let $F/\mathbb{Q}$ be a cyclic cubic extension. If $F/\mathbb{Q}$ is unramified outside $\{q, l\}$ and $q, l$ are ramified in $F/\mathbb{Q}$, then the answer of the problem $P(F, \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^3, 3))$ is affirmative.

This is a direct consequence of Proposition 5.3.

**Theorem 5.5.** Let $L_0/F/\mathbb{Q}$ be a Galois extension satisfying the conditions (C1), (C2), and (C3). Assume that $P(F, \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^3, 3))$ has a solution $L_1$ such that $L_1 \supset L_0$. There is equivalence between

(a) Any prime of $F$ which is ramified in $F/\mathbb{Q}$ is completely decomposed in $L_1/F$;

(b) $P(F, \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^4, 9))$ has a solution $L_2$ such that $L_2 \supset L_1$.

**Proof.** (a) $\Rightarrow$ (b). Let $C$ be the center of $\Gamma(3^4, 9)$, then the order of $C$ is 3 and $\Gamma(3^4, 9)/C$ is isomorphic to $\Gamma(3^3, 3)$. The group $\Gamma(3^4, 9)$ has four maximal
subgroups, one is isomorphic to $\mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ and the others are isomorphic to $\Gamma(3^3, 3)$. Hence there exists a central extension

$$\varepsilon : 1 \to \mathbb{Z}/3\mathbb{Z} \to \Gamma(3^4, 9) \xrightarrow{j} \text{Gal}(L_1/\mathbb{Q}) \to 1$$

such that $j^{-1}(\text{Gal}(L_1/F))$ is isomorphic to $\mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$. The explicit construction of $\varepsilon$ is as follows. We recall that

$$\Gamma(3^4, 9) = \langle x, y, z \mid y[z, x], x^9, y^3, z^3, x^3[y, z], [x, y] \rangle,$$

$$\Gamma(3^3, 3) = \langle a, b, c \mid a^3, b^3, c^3, c[b,a], [a,c], [b,c] \rangle.$$

We can assume that $\text{Gal}(L_1/F) = \langle a, c \rangle$. Indeed maximal subgroups of $\Gamma(3^3, 3)$ are $\langle a, c \rangle$, $\langle ba, c \rangle$, $\langle b^2a, c \rangle$ and $\langle b, c \rangle$. If we replace $ba$ (or $b^2a$) by $a$, then $a, b, c$ satisfy the same relations. And if we replace $b$ by $a$ and $a$ by $b^{-1}$, then $a, b, c$ also satisfy the same relations. Then $j$ is defined by $x \mapsto a$, $y \mapsto c$, $z \mapsto b$.

We shall consider the embedding problem $(L_1/\mathbb{Q}, \varepsilon)$. Let $q$ be a prime of $\mathbb{Q}$ ramified in $L_1/\mathbb{Q}$, and let $\tilde{q}$ be an extension of $q$ to $L_0$. Then $\text{Gal}(L_1q/\mathbb{Q}q)$ is isomorphic to the decomposition group of $\tilde{q}$ in $L_1/\mathbb{Q}$. Since $L_1/F$ is unramified and $\tilde{q}$ is completely decomposed in $L_1/F$, $\text{Gal}(L_1q/\mathbb{Q}q)$ is the cyclic group of order 3 and is not contained in $\text{Gal}(L_1/F)$. Thus $j^{-1}(\text{Gal}(L_1q/\mathbb{Q}q))$ is a subgroup of $\Gamma(3^3, 3)$. Hence the group extension

$$\varepsilon_q : 1 \to \mathbb{Z}/3\mathbb{Z} \to j^{-1}(\text{Gal}(L_1q/\mathbb{Q}q)) \xrightarrow{j} \text{Gal}(L_1q/\mathbb{Q}q) \to 1$$

is split because the exponent of $\Gamma(3^3, 3)$ is 3. In view of Lemma 3.1, the proof of (a) $\Rightarrow$ (b) is complete.

(b) $\Rightarrow$ (a). Let $q$ be a prime of $\mathbb{Q}$ ramified in $F/\mathbb{Q}$, and let $F'$ be the decomposition field of $q$ in $L_0/\mathbb{Q}$. Then $F'$ is a cubic field not equal to $F$. Since $\text{Gal}(L_2/F)$ is isomorphic to $\mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ and other maximal subgroups of $\Gamma(3^4, 9)$ are isomorphic to $\Gamma(3^3, 3)$, $\text{Gal}(L_2/F')$ is isomorphic to $\Gamma(3^3, 3)$. Let $\tilde{q}$ be a prime of $L_0$ lying above $q$. By Lemma 4.1, $\tilde{q}$ is completely decomposed in $L_1/L_0$.

**Theorem 5.6.** Let $L_0/F/\mathbb{Q}$ be a Galois extension satisfying the conditions (C1), (C2), (C3), and (C4). Assume that $P(F, \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^3, 3))$ has a solution $L_1$ such that $L_1 \supset L_0$. There is equivalence between

(a) Any prime of $F$ which is ramified in $F/\mathbb{Q}$ is completely decomposed in $L_1/F$; 
(b) $P(F, \Gamma(3^3, 3), \Gamma(3^4, 9))$ has a solution $L_2$ such that $L_2 \supset L_1$.

**Proof.** Since the proof is similar to that of Theorem 5.5, we merely sketch it. We consider (a) $\Rightarrow$ (b). Let $F'/\mathbb{Q}$ be the cyclic cubic extension as in condition (C4). Then there exists a central extension $\varepsilon : 1 \to \mathbb{Z}/3\mathbb{Z} \to \Gamma(3^4, 9) \xrightarrow{j} \text{Gal}(L_1/\mathbb{Q}) \to 1$ such that $j^{-1}(\text{Gal}(L_1/F)) \cong \Gamma(3^3, 3)$ and that $j^{-1}(\text{Gal}(L_1/F')) \cong \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$.
An application of Lemma 3.1 completes the proof of (a) \(\Rightarrow\) (b). We omit the proof of the converse.

**Theorem 5.7.** Let \(L_0/F/\mathbb{Q}\) be a Galois extension satisfying the conditions (C1), (C2), (C3), and (C4). Assume that \(P(F, \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^3, 3))\) has a solution \(L_1\) such that \(L_1 \supset L_0\). If any prime of \(F\) which is ramified in \(F/\mathbb{Q}\) is completely decomposed in \(L_1/F\), then \(P(F, \Gamma(3^4, 4), \Gamma(3^7, 7))\) has a solution \(L_2\) such that \(L_2 \supset L_1\).

**Proof.** Let \(F'/\mathbb{Q}\) be the cyclic cubic extension as in condition (C4). The maximal subgroups of \(\Gamma(3^4, 7)\) are \(\Gamma(3^3, 3), \Gamma(3^3, 4), \Gamma(3^3, 4), \text{ and } \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}\). Then there exists a central extension

\[\varepsilon : 1 \rightarrow \mathbb{Z}/3\mathbb{Z} \rightarrow \Gamma(3^4, 7) \xrightarrow{j} \text{Gal}(L_1/\mathbb{Q}) \rightarrow 1\]

such that \(j^{-1}(\text{Gal}(L_1/F)) \cong j^{-1}(\text{Gal}(L_1/F')) \cong \Gamma(3^3, 4)\). The explicit construction of \(\varepsilon\) is as follows. We recall that

\[\Gamma(3^4, 7) = \langle x, y, z \mid y[z, x], x^0, y^3, z^3, x^3[y, x], [y, z]\rangle,\]
\[\Gamma(3^3, 3) = \langle a, b, c \mid a^3, b^3, c^3, c[b, a], [a, c], [b, c]\rangle.\]

Here we can assume that \(\text{Gal}(L_1/F) = \langle a, c\rangle\) and \(\text{Gal}(L_1/F') = \langle ab, c\rangle\). Let \(j\) is the group homomorphism defined by \(x \mapsto a, y \mapsto c, z \mapsto b\), then \(j^{-1}(\text{Gal}(L_1/F)) = \langle x, y\rangle \cong \Gamma(3^3, 4)\) and \(j^{-1}(\text{Gal}(L_1/F')) = \langle xz, y\rangle \cong \Gamma(3^3, 4)\).

If \(q\) is a prime of \(\mathbb{Q}\) which is ramified in \(L_1/\mathbb{Q}\), then \(\text{Gal}(L_{1q}/\mathbb{Q}_q)\) is the cyclic group of order 3. Since \(j^{-1}(\text{Gal}(L_{1q}/\mathbb{Q}_q))\) is contained in \(\Gamma(3^3, 3)\) or \(\Gamma(3^3, 5)\), the exponent of \(j^{-1}(\text{Gal}(L_{1q}/\mathbb{Q}_q))\) is equal to 3. Then the group extension

\[\varepsilon_q : 1 \rightarrow \mathbb{Z}/3\mathbb{Z} \rightarrow j^{-1}(\text{Gal}(L_{1q}/\mathbb{Q}_q)) \xrightarrow{j} \text{Gal}(L_{1q}/\mathbb{Q}_q) \rightarrow 1\]

is split. By virtue of Lemma 3.1, the proof is complete. \(\square\)

### 6. Unramified extensions of degree 81 over cyclic cubic fields

Let \(F/\mathbb{Q}\) be a cyclic cubic extension. We consider the case of a Galois extension \(L_3/F/\mathbb{Q}\) such that \(L_3/F\) is unramified extension of degree 81, and the 3-rank of \(\text{Gal}(L_3/\mathbb{Q})\) is 2.

Under these conditions \(\text{Gal}(L_3/\mathbb{Q})\) is isomorphic to one of \(\Gamma(3^5, 3), \Gamma(3^5, 26),\) or \(\Gamma(3^5, 28)\).

In this section we always assume that \(L_0/F/\mathbb{Q}\) satisfies conditions (C1), (C2), (C3), and (C4). Let \(F'\) be the cubic field as in condition (C4).
Theorem 6.1. Assume that the problem \( P(F, \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^4, 9)) \) has a solution \( L_2 \) such that \( L_2 \supset L_0 \). The following conditions are equivalent.

(a) Any prime of \( F \) which is ramified in \( F/\mathbb{Q} \) is completely decomposed in \( L_2/F \).

(b) \( P(F, \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^5, 26)) \) has a solution \( L_3 \) such that \( L_3 \supset L_2 \).

(c) \( P(F, \Gamma(3^4, 4), \Gamma(3^5, 28)) \) has a solution \( L_3 \) such that \( L_3 \supset L_2 \).

Lemma 6.2. Let \( F, F' \) and \( L_0 \) be as in condition (C4). Let \( L_2 \) be a solution of \( P(F, \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^4, 9)) \) such that \( L_2 \supset L_0 \), and let \( L_3/L_2/\mathbb{Q} \) be a Galois extension such that \( L_3/F \) and \( L_3/F' \) are unramified.

(1) If \( \text{Gal}(L_3/\mathbb{Q}) \) is isomorphic to \( \Gamma(3^5, 26) \), we have the equivalence
\[
\text{Gal}(L_3/F) \cong \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z} \iff \text{Gal}(L_3/F') \cong \Gamma(3^4, 9).
\]

(2) If \( \text{Gal}(L_3/\mathbb{Q}) \) is isomorphic to \( \Gamma(3^5, 28) \), we have the equivalence
\[
\text{Gal}(L_3/F) \cong \Gamma(3^4, 4) \iff \text{Gal}(L_3/F') \cong \Gamma(3^4, 10).
\]

Proof. (1) Since one of the maximal subgroups of \( \Gamma(3^5, 26) \) is isomorphic to \( \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z} \) and the others are isomorphic to \( \Gamma(3^4, 9) \), the forward implication is trivial. We consider the reverse implication. Assume that \( \text{Gal}(L_3/F) \cong \Gamma(3^4, 9) \). Let \( F'' \) be the subfield of \( L_3 \) corresponding to the subgroup \( \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z} \). Then \( L_0/F'' \) is not unramified because \( F'' \) is not equal to \( F \) and \( F' \). Since \( \text{Gal}(L_3/F'') \cong \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z} \), there exists a cyclic extension \( M/F'' \) of degree 9 such that \( L_3 \supset M \supset L_0 \). Since \( L_0/F'' \) is not unramified, \( M/L_0 \) is not also unramified. This contradicts that \( L_3/L_0 \) is unramified.

(2) We prove only the forward implication; the converse is similar. Assume that \( \text{Gal}(L_3/F') \) is not isomorphic to \( \Gamma(3^4, 10) \). Let \( F'' \) be the subfield of \( L_3 \) corresponding to the subgroup \( \Gamma(3^4, 10) \), then \( L_0/F'' \) is not unramified. Let \( q \) be a prime of \( F'' \) which is ramified in \( L_0/F'' \) and \( \tilde{q} \) an extension of \( q \) to \( L_2 \). Let \( T \) be the inertia field of \( \tilde{q} \) in \( L_2/F'' \) and \( k \) the intersection of \( L_1 \) and \( T \). Then \( F'' \subseteq k \subseteq T \).
Since $L/H$ from Table 1 that a subgroup $j$ Then $\Gamma_1(T)$. 

Proof of Theorem 6.1. We first consider (a) such that $j$ central extension then the order of $\Gamma_1(T)$ is equal to 3 and the quotient group $\Gamma(3^3, 2)$, the Galois group $Gal(L_3/L_0)$ is isomorphic to $\Gamma(3^3, 2)$. Further one of the maximal subgroups of $Gal(L_3/F''')$ is isomorphic to $\Gamma(3^3, 2)$ and the others are isomorphic to $\Gamma(3^3, 4)$. Then $Gal(L_3/k)$ is isomorphic to $\Gamma(3^3, 4)$. Hence $L_3/T$ is a cyclic extension of degree 9, because one maximal subgroup of $\Gamma(3^3, 4)$ is isomorphic to $Z/3Z \times Z/3Z$ and the others are isomorphic to $Z/9Z$. Since $\hat{q}$ is ramified in $L_2/T$, $\hat{q}$ is also ramified in $L_3/L_2$. This contradicts that $L_3/L_2$ is unramified, proving the desired implication. 

Let $\Gamma(3^3, 28)$ has only one normal subgroup of order 9, which is isomorphic to $Z/3Z \times Z/3Z$. Hence $Gal(L_3/L_1)$ is isomorphic to $Z/3Z \times Z/3Z$. Since all maximal subgroups of $Gal(L_3/F) \cong \Gamma(3^4, 4)$ are isomorphic to $\Gamma(3^3, 2)$, the Galois group $Gal(L_3/L_0)$ is isomorphic to $\Gamma(3^3, 2)$. Further one of the maximal subgroups of $Gal(L_3/F''')$ is isomorphic to $\Gamma(3^3, 2)$ and the others are isomorphic to $\Gamma(3^3, 4)$. Then $Gal(L_3/k)$ is isomorphic to $\Gamma(3^3, 4)$. Hence $L_3/T$ is a cyclic extension of degree 9, because one maximal subgroup of $\Gamma(3^3, 4)$ is isomorphic to $Z/3Z \times Z/3Z$ and the others are isomorphic to $Z/9Z$. Since $\hat{q}$ is ramified in $L_2/T$, $\hat{q}$ is also ramified in $L_3/L_2$. This contradicts that $L_3/L_2$ is unramified, proving the desired implication. 

Proof of Theorem 6.1. We first consider (a) $\Rightarrow$ (b). Let $C$ be the center of $\Gamma(3^5, 26)$, then the order of $C$ is equal to 3 and the quotient group $\Gamma(3^5, 26)/C$ is isomorphic to $\Gamma(3^4, 9)$. The group $\Gamma(3^5, 26)$ has four maximal subgroups, one is isomorphic to $Z/9Z \times Z/9Z$ and the others are isomorphic to $Z/9Z$. Then there exists a central extension

$$\varepsilon : 1 \rightarrow Z/3Z \rightarrow \Gamma(3^5, 26) \rightarrow Gal(L_2/Q) \rightarrow 1$$

such that $j^{-1}(Gal(L_2/F)) \cong Z/9Z \times Z/9Z$ and that $j^{-1}(Gal(L_2/F')) \cong \Gamma(3^4, 9)$. The explicit construction of $\varepsilon$ is as follows. Let $\Gamma(3^5, 26)$ be as on page 171, and

$$\Gamma(3^4, 9) = \langle a, b, c \mid b[c, a], a^0, b^3, c^3, a^3[b, c], [a, b] \rangle.$$ 

By Remark 4.3 we can assume that $Gal(L_2/F) = \langle a, b \rangle$, $Gal(L_2/F') = \langle b, c \rangle$. Then $j$ is defined by $x \mapsto c$, $y \mapsto a$, $z \mapsto b$.

Let $q$ be a prime of $Q$ which is ramified in $L_2/Q$, and $\hat{q}$ an extension of $q$ to $L_2$. Then $Gal(L_2/q/Q_q)$ is isomorphic to the decomposition group of $\hat{q}$ in $L_2/Q$. Since $L_2/F$ is unramified and $\hat{q}$ is completely decomposed in $L_2/F$, $Gal(L_2/q/Q_q)$ is the cyclic group of order 3 and is not contained in $Gal(L_2/F)$. Now, we see from Table 1 that a subgroup $H$ of $\Gamma(3^4, 9)(\cong Gal(L_3/F'))$ having order 27 and
not contained in $\Gamma(3^4, 2)(\cong \text{Gal}(L_3/F))$ must be isomorphic to $\Gamma(3^3, 3)$. Thus $j^{-1}(\text{Gal}(L_{2q}/\mathbb{Q}_q))$ is a subgroup of $\Gamma(3^3, 3)$. Since the exponent of $\Gamma(3^3, 3)$ is 3, the group extension

$$\varepsilon_q : 1 \to \mathbb{Z}/3\mathbb{Z} \to j^{-1}(\text{Gal}(L_{2q}/\mathbb{Q}_q)) \xrightarrow{j} \text{Gal}(L_{2q}/\mathbb{Q}_q) \to 1$$

is split. In view of Lemma 3.1, the embedding problem $(L_2/\mathbb{Q}, \varepsilon)$ has a proper solution $L_3$ such that $L_3/L_2$ is unramified. Since $\text{Gal}(L_3/F)$ is isomorphic to $j^{-1}(\text{Gal}(L_2/F)) = \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z}$, $L_3$ is a required field.

Next we consider (b) $\Rightarrow$ (a). Let $q$ be a prime of $\mathbb{Q}$ which is ramified in $F/\mathbb{Q}$, and $\hat{q}$ an extension of $q$ to $L_2$. Assume that $\hat{q}$ is not completely decomposed in $L_2/F$. Let $L_0$ be the field such that $L_0 \subset L_1 \subset L_2$ and that $\text{Gal}(L_1/\mathbb{Q}) \cong \Gamma(3^3, 3)$. Then by Theorem 5.5 and the assumption, $\hat{q}$ is completely decomposed in $L_1/F$ and is inert in $L_2/L_1$. Let $F''$ be the decomposition field of $q$ in $L_0/\mathbb{Q}$. Let $T$ be the inertia field of $\hat{q}$ in $L_2/\mathbb{Q}$ and $k$ be the intersection of $L_1$ and $T$. Then $F'' \subseteq k \subseteq T$. We refer the field diagram in the proof of Lemma 6.2.

Since $\text{Gal}(L_3/F'')$ is a maximal subgroup of $\text{Gal}(L_3/\mathbb{Q})$ and $\text{Gal}(L_3/F) \cong \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z}$, $\text{Gal}(L_3/F'')$ is isomorphic to $\Gamma(3^4, 9)$. Since $\text{Gal}(L_3/k)$ is a maximal subgroup of $\text{Gal}(L_3/F'')$ and $\text{Gal}(L_3/L_0) \cong \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$, $\text{Gal}(L_3/k)$ is isomorphic to $\Gamma(3^3, 3)$. This contradicts Lemma 4.1.

The proof of (a) $\iff$ (c) is similar to that of (a) $\iff$ (b), so we only sketch it. Consider (a) $\Rightarrow$ (c). There exists a central extension

$$\varepsilon : 1 \to \mathbb{Z}/3\mathbb{Z} \to \Gamma(3^5, 28) \xrightarrow{j} \text{Gal}(L_2/\mathbb{Q}) \to 1$$

such that $j^{-1}(\text{Gal}(L_2/F)) \cong \Gamma(3^4, 4)$ and $j^{-1}(\text{Gal}(L_2/F'')) \cong \Gamma(3^4, 10)$. The explicit construction of $\varepsilon$ is as follows. Let $\Gamma(3^5, 28)$ be as on page 171, and set

$$\Gamma(3^4, 9) = \langle a, b, c | b[c, a], a^9, b^3, c^3, a^3[b, c], [a, b] \rangle.$$
Theorem 6.3. Assume that the problem $P(F, \Gamma(3^3, 3), \Gamma(3^4, 9))$ has a solution $L_2$ such that $L_2 \supset L_0$. The following conditions are equivalent.

(a) Any prime of $F$ which is ramified in $F/\mathbb{Q}$ is completely decomposed in $L_2/F$.
(b) $P(F, \Gamma(3^4, 9), \Gamma(3^5, 26))$ has a solution $L_3$ such that $L_3 \supset L_2$.
(c) $P(F, \Gamma(3^4, 10), \Gamma(3^5, 28))$ has a solution $L_3$ such that $L_3 \supset L_2$.

This follows trivially from Theorem 6.1 and Lemma 6.2.

Proposition 7.1. Let $L, F, F'$ be as above.

(1) The class number of $F$ is divisible by 9 if and only if the class number of $F'$ is divisible by 9. Further in this case, the ideal class group of $F$ and $F'$ has a subgroup $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$.

(2) The class number of $F$ is divisible by 27 if and only if the class number of $F'$ is divisible by 27. Further in this case, the ideal class group of $F$ and $F'$ has a subgroup $\mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$.

(3) The class number of $F$ is divisible by 81 if and only if the answer of the problem $P(F', \Gamma(3^4, 10))$ is affirmative. Further in this case, the ideal class group of $F$ has a subgroup $\mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z}$.

Lemma 7.2. Let $p$ be an odd prime and $F/\mathbb{Q}$ a $p$-extension. If the class number of $F$ is divisible by $p^r$ for some integer $r$, then there exists a Galois extension $M/F/\mathbb{Q}$ such that $M/F$ is unramified abelian and the degree $[M:F]$ is equal to $p^r$. 

\[ \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \text{ Gal}(L_3/k) \text{ is isomorphic to } \Gamma(3^3, 3). \text{ This contradicts Lemma 4.1.} \]
Proof: By class field theory, there exists an unramified abelian extension $K/F$ such that the degree $[K : F]$ is equal to $p'$. Let $M_1/\mathbb{Q}$ be the Galois closure of $K/\mathbb{Q}$. Then $M_1/F/\mathbb{Q}$ is a Galois extension such that $M_1/F$ is unramified abelian $p$-extension and the degree $[M_1 : F]$ is greater than or equal to $p'$. If $[M_1 : F] = p'$ then $M_1/F/\mathbb{Q}$ is a required field. Assume that $[M_1 : F] > p'$. Let $C(\text{Gal}(M_1/\mathbb{Q}))$ be the center of $\text{Gal}(M_1/\mathbb{Q})$. Since $\text{Gal}(M_1/\mathbb{Q})$ is a $p$-group and $\text{Gal}(M_1/F)$ is a normal subgroup of $\text{Gal}(M_1/\mathbb{Q})$, the intersection $\text{Gal}(M_1/F) \cap C(\text{Gal}(M_1/\mathbb{Q}))$ is nontrivial. Then there exists a Galois extension $M_2/F/\mathbb{Q}$ such that $M_2/F$ is unramified $p$-extension and the degree $[M_2 : F]$ is equal to $[M_1 : F]/p$. By repeating this process, we get the required extension $M/F/\mathbb{Q}$. $\square$

Proof of Proposition 7.1. (1) Assume that the class number of $F$ is divisible by 9. By Lemma 7.2 there exists a Galois extension $L_1/F/\mathbb{Q}$ such that $L_1/F$ is unramified abelian and that $[L_1 : F] = 9$. By Lemma 4.2 and the assumption for the number of ramified primes, $\text{Gal}(L_1/\mathbb{Q})$ is generated by two elements of order 3. Then $\text{Gal}(L_1/\mathbb{Q})$ is isomorphic to $\Gamma(3^3, 3)$. Thus $L_1/F^*$ is unramified and $\text{Gal}(L_1/F^*) \cong \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$, because all maximal subgroups of $\Gamma(3^3, 3)$ are isomorphic to $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$. The proof of the converse is similar.

(2) Assume that the class number of $F$ is divisible by 27. By Lemma 7.2 there exists a Galois extension $L_2/F/\mathbb{Q}$ such that $L_2/F$ is unramified abelian and that $[L_2 : F] = 27$. Since $\text{Gal}(L_2/\mathbb{Q})$ is generated by two elements of order 3, $\text{Gal}(L_2/\mathbb{Q})$ is isomorphic to $\Gamma(3^4, 7)$ or $\Gamma(3^4, 9)$. We claim that $\text{Gal}(L_2/\mathbb{Q})$ is not isomorphic to $\Gamma(3^4, 7)$. We assume $\text{Gal}(L_2/\mathbb{Q}) \cong \Gamma(3^4, 7)$. Since $\Gamma(3^4, 7)$ has two maximal subgroups which are isomorphic to $\Gamma(3^3, 4)$, there exists a cubic field $F''$ such that $\text{Gal}(L_2/F'') \cong \Gamma(3^3, 4)$ and that $F'' \neq F, F'$. Then only one prime ramifies in $F''/\mathbb{Q}$. By Iwasawa [Iwasawa 1956] the class number of $F''$ is prime to 3. Since $\Gamma(3^3, 4)$ is not generated by elements of order 3, this contradicts Lemma 4.2. Then $L_2$ is a solution of $P(F, \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^4, 9))$.

Let $C$ be the center of $\Gamma(3^4, 9)$, and $L_1$ the subfield of $L_2$ corresponding to $C$. Since $\Gamma(3^4, 9)/C$ is isomorphic to $\Gamma(3^3, 3)$, $L_1$ is a solution of the problem $P(F, \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^3, 3))$. By Theorem 5.5 any prime of $F$ which is ramified in $F/\mathbb{Q}$ is completely decomposed in $L_1/F$. Since all maximal subgroups of $\Gamma(3^3, 3)$ are isomorphic to $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$, $L_1$ is also a solution of the problem $P(F', \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^3, 3))$. Then $P(F', \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^4, 9))$ has a solution $L_2'$ by Theorem 5.5. Hence the class number of $F'$ is divisible by 27. The proof of the converse is similar.

(3) Assume that the class number of $F$ is divisible by 81. By Lemma 7.2 there exists a Galois extension $L_3/F/\mathbb{Q}$ such that $L_3/F$ is unramified abelian and that $[L_3 : F] = 81$. Since $\text{Gal}(L_3/\mathbb{Q})$ is generated by two elements of order 3, $\text{Gal}(L_3/\mathbb{Q})$ is isomorphic to $\Gamma(3^5, 26)$ and $\text{Gal}(L_3/F)$ is isomorphic to $\mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z}$. Let
Let $C$ be the center of $\Gamma(3^5, 26)$, and $L_2$ the subfield of $L_3$ corresponding to $C$. Since $\Gamma(3^5, 26)/C$ is isomorphic to $\Gamma(3^4, 9)$, $L_2$ is a solution of the problem $P(F, \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^3, 9))$. By Theorem 6.1 any prime of $F$ which is ramified in $F/\mathbb{Q}$ is completely decomposed in $L_2/F$, and $L_2$ is also a solution of $P(F', \Gamma(3^3, 3), \Gamma(3^4, 9))$. By Theorem 6.3 the problem $P(F', \Gamma(3^4, 10), \Gamma(3^5, 28))$ has a solution $L'_3$. 

For the converse we assume that $L_3$ is a solution of $P(F', \Gamma(3^4, 10))$. Then $\Gamma := \text{Gal}(L_3/\mathbb{Q})$ has order $243$ and 3-rank $2$, and it has a maximal subgroup isomorphic to $\Gamma(3^4, 10)$. The group satisfying these conditions is isomorphic to $\Gamma(3^5, 28)$. We prove that the Galois group $\text{Gal}(L_3/F)$ is isomorphic to $\Gamma(3^4, 4)$, which is a maximal subgroup of $\Gamma(3^5, 28)$. For the proof, we assume that $\text{Gal}(L_3/F)$ is not isomorphic to $\Gamma(3^4, 4)$, and let $F''$ be the subfield of $L_3$ corresponding to $\Gamma(3^4, 4)$. Then $F''$ is not equal to $F$ and $F'$. Hence, by [Iwasawa 1956], the class number of $F''$ is prime to 3. By Lemma 4.2, $\text{Gal}(L_3/F'')$ must be generated by elements of order 3. But $\Gamma(3^4, 4)$ is not generated by elements of order 3. This is a contradiction.

Let $L_2$ be the subfield of $L_3$ corresponding to the center of $\text{Gal}(L_3/\mathbb{Q})$, then $\text{Gal}(L_2/\mathbb{Q}) \cong \Gamma(3^4, 9)$. Let $C$ be the center of $\text{Gal}(L_3/F)$. Since $\text{Gal}(L_3/F) \cong \Gamma(3^4, 4)$ and $\text{Gal}(L_2/F) \cong \text{Gal}(L_3/F)/C \cong \Gamma(3^3, 2)$, then $\text{Gal}(L_2/F') \cong \Gamma(3^3, 3), \Gamma(3^4, 9))$. By Theorem 6.3, any prime of $F'$ which is ramified in $F'/\mathbb{Q}$ is completely decomposed in $L_2/F'$. Hence any prime of $F$ which is ramified in $F/\mathbb{Q}$ is completely decomposed in $L_2/F$. $L_2$ is also a solution of the problem $P(F, \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}, \Gamma(3^4, 9))$. By Theorem 6.1, the problem $P(F, \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z}, \Gamma(3^5, 26))$ has a solution $L'_3$. Then $L'_3/F$ is unramified abelian extension and the Galois group is isomorphic to $\mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z}$. □

**Example 7.3.** Let $F_{pq}$ and $F'_{pq}$ denote the two cyclic cubic fields of conductor $pq$, where $p \equiv q \equiv 1 \pmod{3}$, and let $L = F_{pq}F'_{pq}$ be their composite. Denote by $(n_1, n_2, \ldots, n_r)$ the group $\mathbb{Z}/n_1\mathbb{Z} \times \mathbb{Z}/n_2\mathbb{Z} \times \cdots \times \mathbb{Z}/n_r\mathbb{Z}$. The following table contains a few class groups computed with PARI:

<table>
<thead>
<tr>
<th>$p$</th>
<th>$q$</th>
<th>$\text{Cl}(F_{pq})$</th>
<th>$\text{Cl}(F'_{pq})$</th>
<th>$\text{Cl}(L)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>181</td>
<td>(6,6)</td>
<td>(3,3)</td>
<td>(6,2)</td>
</tr>
<tr>
<td>43</td>
<td>193</td>
<td>(3,3)</td>
<td>(3,3)</td>
<td>(3,3)</td>
</tr>
<tr>
<td>73</td>
<td>241</td>
<td>(9,3)</td>
<td>(63,3)</td>
<td>(21,3,3)</td>
</tr>
<tr>
<td>79</td>
<td>157</td>
<td>(9,3)</td>
<td>(9,3)</td>
<td>(9,3,3)</td>
</tr>
<tr>
<td>181</td>
<td>331</td>
<td>(9,3)</td>
<td>(9,3)</td>
<td>(3,3,3,3)</td>
</tr>
<tr>
<td>103</td>
<td>409</td>
<td>(9,9)</td>
<td>(27,9)</td>
<td>(9,9,3,3)</td>
</tr>
</tbody>
</table>
Corollary 7.4. Let $L, F, F'$ be as above.

(1) Assume that the class number of $F$ is divisible by 27. Then the problem
$P(F, \Gamma(3^4, 3), \Gamma(3^5, 3))$ has a solution. In particular the class number of
the Hilbert 3-class field of $F$ is divisible by 3.

(2) Assume that the class number of $F$ is divisible by 81. Then the problem
$P(F, \Gamma(3^5, 2), \Gamma(3^6, 40))$ has a solution.

Proof. (1) Let $L_1, L_2, L_2'$ be as in the proof of Proposition 7.1(2). By the proof of
Proposition 7.1(2), $\text{Gal}(L_2/\mathbb{Q})$ is not isomorphic to $\text{Gal}(L_2'/\mathbb{Q})$. Then $L_2 \neq L_2'$.
Let $\bar{L}$ be the composition field of $L_2$ and $L_2'$. Since $\text{Gal}(\bar{L}/L_2)$ and $\text{Gal}(\bar{L}/L_2')$ are
contained in the center of $\text{Gal}(\bar{L}/\mathbb{Q})$, then the center has a subgroup isomorphic to
$\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$. In addition, $\text{Gal}(\bar{L}/\mathbb{Q})$ has order 243, has 3-rank 2, and is generated
by elements of order 3. The group satisfying these conditions is isomorphic to
$\Gamma(3^5, 3)$. $\Gamma(3^5, 3)$ has four maximal subgroups, two are isomorphic to $\Gamma(3^4, 3)$ and
the others are isomorphic to $\Gamma(3^5, 12)$. We remark that $\Gamma(3^4, 3)$ is not generated
by elements of order 3. Let $F''$ and $F'''$ are cyclic cubic subfield of $\bar{L}$ not equal
to $F$ and $F'$. Then by Iwasawa [1956], the class number of $F''$ and $F'''$ are both
prime to 3. Since $\text{Gal}(\bar{L}/F'')$ and $\text{Gal}(\bar{L}/F''')$ are generated by elements of order 3,
$\text{Gal}(\bar{L}/F'') \cong \text{Gal}(\bar{L}/F''') \cong \Gamma(3^4, 12)$. Hence $\text{Gal}(\bar{L}/F) \cong \text{Gal}(\bar{L}/F') \cong \Gamma(3^4, 3)$.

(2) Let $L_3, L_3', L_3''$ be as in the proof of Proposition 7.1(3). By that same proof
we have $L_3 \neq L_3'$. Let $\bar{L}$ be the composite of $L_3$ and $L_3'$. Then $\text{Gal}(\bar{L}/\mathbb{Q})$ has
order 243 and 3-rank 2, it is generated by elements of order 3, and its center has
a subgroup isomorphic to $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$. The group satisfying these conditions
is isomorphic to $\Gamma(3^5, 53)$ and the others are isomorphic to $\Gamma(3^5, 2)$ or $\Gamma(3^5, 15)$. We
remark that $\Gamma(3^5, 2)$ and $\Gamma(3^5, 15)$ are not generated by elements of order 3. Then
$\text{Gal}(\bar{L}/F)$ is isomorphic to $\Gamma(3^5, 2)$ or $\Gamma(3^5, 15)$. Since $\Gamma(3^5, 15)$ has no subgroup
$H$ such that $\Gamma(3^5, 15)/H \cong \text{Gal}(L_3/F)(\cong \mathbb{Z}/9\mathbb{Z} \times \mathbb{Z}/9\mathbb{Z})$, $\text{Gal}(\bar{L}/F)$ is isomorphic
to $\Gamma'(3^5, 2)$.

Acknowledgment

I am grateful to the referee for a careful reading, for advice, and for supplying
Example 7.3.

References

[Boston and Leedham-Green 1999] N. Boston and C. Leedham-Green, “Counterexamples to a con-
11095

[Cohn 1978] H. Cohn, A classical invitation to algebraic numbers and class fields, Springer, New
York, 1978. MR 80c:12001 Zbl 0395.12001


Received August 7, 2005. Revised November 3, 2005.

AKITO NOMURA

DEPARTMENT OF MATHEMATICS

KANAZAWA UNIVERSITY

KANAZAWA 920-1192

JAPAN

anomura@t.kanazawa-u.ac.jp
We consider generalizations of Shanks’ sequence of quadratic fields \( \mathbb{Q}(\sqrt{S_n}) \) where \( S_n = (2^n + 1)^2 + 2^{n+2} \). Quadratic fields of this type are of interest because it is possible to explicitly determine the fundamental unit. If a sequence of quadratic fields given by \( D_n = A^2 x^{2n} + B x^n + C^2 \) satisfies certain conditions (notably that the regulator is of order \( \Theta(n^2) \)), then we determine the exact form such a sequence must take.

1. History of creepers

We will be interested in simple continued fractions, which we denote by \( \alpha = [a_0, a_1, a_2, \ldots] \); the \( a_i \) are called the partial quotients of \( \alpha \). It is well-known that a continued fraction expansion is periodic if and only if it is the expansion of a real quadratic irrational. We denote the period length of a real quadratic irrational \( \alpha \) by \( lp(\alpha) \).

For real quadratic fields, it is expected that the class number will usually be small; see [Cohen and Lenstra 1984]. By the correspondence between ideals and continued fractions this is equivalent to the continued fraction expansion of \( \sqrt{D} \) being long, generally of length about \( \sqrt{D} \). Thus, examples of short expansions of \( \sqrt{D} \) should be considered as unusual and worthy of interest.

It is easy to find sequences of integers \( D_i \) such that \( \sqrt{D_i} \) has a bounded period length. Many results have been determined for such families and we refer the reader to [Perron 1950; van der Poorten and Williams 1999; Schinzel 1960; 1961].

Shanks [1969] examined the class numbers of quadratic fields with discriminants given by \( n^2 - 2^{2k+1} \). He noticed that for the family \( S_n = (2^n + 3)^2 - 8 \), the class number of \( S_n \) grows infinitely large. This sequence of fields is known as Shanks’ sequence. It happens that Shanks’ sequence is just a special case of an earlier
example given in [Nyberg 1949]:

\[(1-1) \quad D_n = (x^n + (x \pm 1)/2)^2 \mp x^n.\]

The ring of algebraic integers of a quadratic number field \(K = \mathbb{Q}(\sqrt{D})\), denoted by \(\mathcal{O}_K\), is equal to \(\mathbb{Z}[\omega']\), where \(d_K\) is the squarefree kernel of \(D\) and

\[\omega' = \begin{cases} \sqrt{d_K} & \text{if } d_K \not\equiv 1 \pmod{4}, \\ (1 + \sqrt{d_K})/2 & \text{if } d_K \equiv 1 \pmod{4}. \end{cases}\]

The discriminant, \(D_K\), of \(\mathcal{O}_K\) is equal to \(d_K\) if \(d_K\) is congruent to 1 modulo 4 and \(4d_K\) otherwise. An order \(\mathcal{O}\) of \(K\) is defined to be a subring of \(K\), containing 1, such that the quotient \(\mathcal{O}/\mathbb{Z}\) is finite, such an order must be of the form \(\mathcal{O} = \mathbb{Z}[f\omega']\). The number \(f\) is called the conductor of \(\mathcal{O}\). The discriminant of an order \(\mathcal{O} \subset \mathcal{O}_K\) is equal to \(D = f^2D_k\). Thus, the discriminant of an order is always congruent to 0 or 1 modulo 4. The discriminant of the maximal order is called a fundamental discriminant.

For any discriminant, \(D \equiv t \pmod{4}\), the element \(\omega = (t + \sqrt{D})/2\) is an algebraic integer since \(t \equiv 0, 1 \pmod{4}\). If we know that \(D\) is fundamental then we usually write \(\omega'\) instead of \(\omega\). With this notation, the expansion of \(\omega_n = (1+\sqrt{S_n})/2\) corresponding to Shanks’ sequence has a period length of \(2n+1\),

\[\omega_n = [2^{n-1} + 1, 1, 2^{n-1}, 2, 2^{n-2}, 2^2, \ldots, 2^{n-1}, 1, 2^n + 1].\]

The fundamental unit of the order with discriminant \(D_n\) is given by

\[\varepsilon_n = \left((2^n + 1 + \sqrt{D_n})/2\right)(2^n + 3 + \sqrt{D_n})/4)^n.\]

The regulator of the order \(\mathcal{O}\), denoted by \(R(\mathcal{O})\) or by \(R(D)\) if \(\mathcal{O}\) has discriminant \(D\), is defined as the logarithm of the fundamental unit. Thus, sequences of discriminants \(D_n\), where \(\omega_n\) has a bounded period length have regulators of order \(O(\log D_n)\). Examples like Shanks’ sequence have regulators of order \(O((\log D_n)^2)\).

Several people have since generalized Shanks’ sequence. They include Hendy [1974], Bernstein [1976a; 1976b; 1976c], Azuhata [1984; 1987], and Levesque and Rhin [1986]. A more synthetic account was given in [Williams 1985]. The most general form was presented in [Williams 1995] as

\[D_n = (qr x^n + \mu(x^k - \lambda)/q)^2 + 4\lambda r x^n,\]

with \(\mu, \lambda \in \{-1, 1\}\) and \(rq |x^k - \lambda\). The automata of Raney were used in [van der Poorten 1994] to provide an alternate way of constructing \(\omega_n\) and \(\varepsilon_n\).
Kaplansky [1998] coined the terms “sleepers”, a sequence of discriminants whose period lengths are bounded, “creepers”, a sequence whose lengths gently go to infinity, and “leapers”, the generic discriminants whose period lengths increase exponentially.

By selecting a sequence of discriminants from families of sleepers appropriately, one can form a sequence of discriminants with linear period length. These are known as “beepers” and can be found in [Mollin and Cheng 2002; van der Poorten 1999; Williams and Buck 1994]. Since these discriminants are selected from sleepers they have a regulator of order $O(\log D_n)$.

These two ideas were used simultaneously by Madden [2001], who explicitly constructed a sequence of discriminants whose continued fraction expansions possessed slowly growing period lengths. These examples were distinct from the known creepers since they were not polynomially parametrized. However, they can be viewed as selecting specific discriminants from various families we will construct here, much as beepers are specially selected sleepers.

We define a creeper to be an infinite family of discriminants $D_n$, such that $f(X, n) \in \mathbb{Q}[X, X^n]$ and for a fixed $x \in \mathbb{Q}$ we have $D_n = f(x, n)$ satisfying

$$lp(\omega_n) = an + b \quad \text{with} \quad a, b \in \mathbb{Q} \quad \text{and} \quad R(D_n) = \Theta(n^2).$$

Kaplansky [1998] made several conjectures about creepers which are quadratic in $x^n$. He suggested that every such creeper could be written as $D_n = A^2 x^{2n} + Bx^n + C^2$ with $A, B, C \in \mathbb{Q}$. Each of the examples upon which these conjectures were based has a principal ideal whose norm is a fixed power of $x$. Consequently, we define a kreeper to be an infinite sequence of discriminants $D_n$ such that

1. $D_n = A^2 x^{2n} + Bx^n + C^2$, where $A, B, C \in \mathbb{Q}$, and $x \in \mathbb{Z}^+.$
2. $lp(\omega_n) = an + b$, where $a$ and $b$ are rational numbers.
3. In the principal cycle there exists an element whose norm is $x^g$ for some $g$ fixed independently of $n$.

Note that the existence of some $Q_b = x^g$ implies $R(D_n) = \Theta(n^2)$. In other words, every kreeper is also a creeper. A proof of this is given in [Patterson 2003, Theorem 17]. Indeed many details are excluded here, and can be found in the same reference.

The main results here are the following.

**Theorem 1.1.** Any kreeper $D_n$ can be written as

$$d^2 D_n = c^2 \left( (qr x^n + (mz^2 x^k - l y^2)/q)^2 + 4 l y^2 r x^n \right).$$

---

1 By “gently” he meant that the periods could be written in an arithmetic progression involving a parameter $n$ used in the presentation of the family.
where each term in the above equation is an element of \( \mathbb{Z} \), the terms \( r, l, m \) are squarefree, \( r, x \) are positive, and the following conditions hold:

\[
(qr x, mlzy) = 1, \quad (qr, x) = 1, \quad (mz, ly) = 1, \\
q | mz^2 x^k - ly^2, \quad c^2 r ly^2 mz^2 | d^2 D_n.
\]

**Theorem 1.2.** Any sequence of discriminants given by (1-2) and satisfying the conditions (1-3) as above must in fact be a kreeper.

As a final introductory remark, we mention that Shanks’ sequence has also been generalised to certain cubic fields with unit rank one, see [Adam 1995; 1998; Levesque and Rhin 1991].

### 2. Preliminary results

Removing nonpositive partial quotients from a continued fraction is not difficult, however removing a fractional partial quotient is, in general, quite difficult. The following few results do provide some assistance. Multiplication can be accomplished via

\[
x[a, b, c, d, \ldots] = [ax, b/x, xc, d/x, \ldots],
\]

which leads to:

**Lemma 2.1** (Folding Lemma [Mendès France 1973]). Let \( x/y = [a_0, a_1, \ldots, a_h] \) with \( (x, y) = 1 \), and denote the sequence \( a_1, \ldots, a_h \) by \( \overrightarrow{w} \) (where \( \overrightarrow{w} \) corresponds to the sequence \( a_h, \ldots, a_1 \)). Then

\[
\frac{x}{y} + \gamma = \left[ a_0, \overrightarrow{w}, c - y'/y \right] = \left[ a_0, \overrightarrow{w}, c, -\overrightarrow{w} \right],
\]

where \( y/y' = [a_h, \ldots, a_1] \), \( (y, y') = 1, y' > 0 \).

This result is more than just a novelty. Besides our use of it here, in [van der Poorten 2002] it is used to rediscover the symmetry formulas.

A result which will be pivotal to our expansions later on is the following simple lemma.

**Lemma 2.2.** If \( x/y = [a_0, \overrightarrow{w}] \), where \( \overrightarrow{w} \) is defined as above then

\[
x/y + \gamma = \left[ a_0, \overrightarrow{w}, \frac{(-1)^h}{y} y^2 - b \right],
\]

where \( b \) is equal to \( (-1)^{h+1}/x \) modulo \( y \).

**Proof.** Using the Folding Lemma we obtain,

\[
\frac{x}{y} + \gamma = \frac{x}{y} + \frac{(-1)^h}{y^2}(-1)^h y y^2 = \left[ a_0, \overrightarrow{w}, \frac{(-1)^h}{y} y^2 - b \right],
\]
where \( b \) satisfies \( xb - cy = (-1)^{h+1} \), which implies \( b \equiv (-1)^{h+1}/x \pmod{y} \).

If the minimal polynomial of \( \alpha \) is \( x^2 - tx + n \) then a typical line in the continued fraction expansion of \( \alpha \) appears as

\[
\frac{\alpha + P_h}{Q_h} = a_h - \frac{(\overline{\alpha} + P_{h+1})}{Q_h},
\]

where \( P_0 = 0, Q_0 = 1 \) and \( \overline{\alpha} \) represents the nontrivial automorphism of the quadratic number field. We then have

\[
P_{h+1} = a_h Q_h - P_h - t \quad \text{and} \quad Q_{h+1} = -\frac{n + P_h(P_h + t)}{Q_h}
\]

A quadratic irrational \( \alpha \) is called reduced if \( \alpha > 1 \) and \( -1 < \overline{\alpha} < 0 \). An integral ideal \( \mathfrak{a} \) is called primitive if it is not divisible by any element of \( \mathbb{Z} \). An integral ideal \( \mathfrak{a} \) is reduced if there does not exist any nonzero \( \alpha \in \mathfrak{a} \) satisfying

\[
|\alpha| < N(\mathfrak{a}) \quad \text{and} \quad |\overline{\alpha}| < N(\mathfrak{a}).
\]

If \( \mathfrak{a} \) is a primitive ideal such that \( N(\mathfrak{a}) < \sqrt{D}/2 \) then \( \mathfrak{a} \) is reduced. From now on, ideal will mean “integral ideal”.

One of the uses of the continued fraction expansion of a quadratic irrational is the determination of the fundamental unit. Rather than keeping track of the convergents, this can be done via the following result.

**Proposition 2.3.** If \( \alpha = [a_0, a_1, \ldots, a_i, \alpha_{i+1}] \) and \( x_j/y_j = [a_0, \ldots, a_j] \), where \( (x_j, y_j) = 1 \), then

\[
\alpha_1 \alpha_2 \ldots \alpha_{h+1} = (-1)^{h+1}(x_h - y_h \overline{\alpha})^{-1}.
\]

**Corollary 2.4.** If \( \mathcal{O} \) is an order of \( \mathbb{Q}(\sqrt{D}) \) and \( \alpha_i, 1 \leq i \leq h+1 \) represents a system of reduced elements in any cycle of quadratic irrationals in \( \mathcal{O} \), then \( \varepsilon = \prod_{i=1}^{h+1} \alpha_i \) is the fundamental unit of \( \mathcal{O} \).

Such a cycle of quadratic irrationals is produced by the continued fraction expansion. To be more precise, if \( a_0 = Q_0 \mathbb{Z} + (P_0 + \omega) \mathbb{Z} \) is an ideal of \( \mathcal{O} \) then the continued fraction expansion of \( (\omega + P_0)/Q_0 \) produces a sequence of complete quotients \( (\omega + P_i)/Q_i \) such that the ideals associated to each complete quotient, that is \( \mathfrak{a}_i = Q_i \mathbb{Z} + (\omega + P_i) \mathbb{Z} \), are all equivalent to \( \mathfrak{a}_0 \).

Later we will need to transfer results from one order to another, where the following proposition will be useful.

---

\(^2\)The \( P_h, Q_h \) appearing here are not, in general, the same as those used in [Perron 1950]
Proposition 2.5. Let \( \mathcal{O}_1, \mathcal{O}_2 \) be two orders of a real quadratic field given by \( \mathcal{O}_1 = \mathbb{Z}[f \omega'], \mathcal{O}_2 = \mathbb{Z}[g \omega'] \) then
\[
R(\mathcal{O}_1) > \frac{(g, f)}{2g} R(\mathcal{O}_2).
\]

Let \( \alpha \) be any ideal of the order \( \mathcal{O} \) having conductor \( f \) in \( K \), and suppose that \( (N(\alpha), f) = 1 \). Then \( \alpha = (t)rs \), where \( t \in \mathbb{Z} \) and any prime ideal divisor of \( r \) lies over a prime which ramifies and any prime ideal divisor of \( s \) lies over a prime which splits in \( \mathcal{O} \). Furthermore, \( r \) and \( s \) are primitive. We will denote by \( s(\alpha) \) the ideal \( s \). Note that if \( t = 1 \) (in which case \( \alpha \) is primitive) then \( \alpha^2 = (r)s^2 \), where \( r \) is squarefree and \( r \mid D \). Also note that \( N(s(\alpha)) = N(s(\tilde{\alpha})) \).

We now introduce a generalisation of a result of Yamamoto [1971].

Definition 1. Let \( b_1, \ldots, b_n \) be invertible ideals of an order \( \mathcal{O} \). We say that \( b_1, \ldots, b_n \) are independent in \( \mathcal{O} \) if whenever there exist nonzero integers \( u, v \) and \( \alpha_i, \beta_i \in \mathbb{Z}^+ (i = 1, \ldots, n) \) such that
\[
(u) \prod_{i=1}^{n} b_i^{\alpha_i} = (v) \prod_{i=1}^{n} b_i^{\beta_i}
\]
then \( \alpha_i = \beta_i (i = 1, \ldots, n) \).

A sufficient condition for the independence of two ideals \( b \) and \( c \) in \( \mathcal{O} \) is given in Theorem 2.7, which needs the following Lemma.

Lemma 2.6. If \( b \) and \( c \) are dependent in \( \mathcal{O} \) then there exist nonzero integers \( u, v \) and nonnegative integers \( m, n \), with \( m + n > 0 \), such that
\[
(u)b^m = (v)\mathfrak{d}^n,
\]
where \( \mathfrak{d} \) is equal to \( c \) or \( \bar{c} \).

Theorem 2.7. If \( b \) and \( c \) are dependent in \( \mathcal{O} \) and \( (N(b)N(c), f) = 1 \) then for some nonnegative integers \( m, n \), with \( m + n > 0 \) we have
\[
N(s(b))^m = N(s(c))^n.
\]

Proof. By the preceding Lemma, we know there must exist integers \( m, n, u, v \) such that \( (u)b^m = (v)\mathfrak{d}^n \), where \( m, n \) are nonnegative and at least one of \( m \) or \( n \) is positive. Then we can write \( (ut_1^m)^{\tilde{b}^m} = (vt_2^n)^{\tilde{d}^n} \), where \( \tilde{b} \) and \( \tilde{d} \) are primitive.

The condition \( (f, N(b)N(c)) = 1 \) and primitivity allow us to write
\[
(ut_1^m)^{\tilde{b}^m} = (vt_2^n)^{r(b)^m s(b)^n} \quad \text{and} \quad (ut_2^n)^{\tilde{d}^n} = (vt_1^m)^{r((d))^n s((d))^n}.
\]

Squaring these yields
\[
(u^2t_1^m t_1^{2m})s(b)^{2m} = (v^2t_2^n t_2^{2n})s((d))^{2n},
\]
where \( r_1 \) and \( r_2 \) divide \( D_n \). Dividing out any common factors of \( u^2 r_1^{m_1} t_1^{2m} \) and \( v^2 r_2^{m_2} t_2^{2n} \) provides coprime integers \( u_1, v_1 \) such that

\[
(u_1) \cdot (s(b))^{2m} = (v_1) \cdot (s(d))^{2n}.
\]

Let \( \mathcal{P} \) and \( \mathcal{Q} \) be a set of discriminants of the form \( D_i \). If \( \mathcal{P} \) contains \( D_i \) and \( \mathcal{Q} \) does not contain \( D_i \) then \( N(\mathcal{P}) \) implies that \( \mathcal{P} \) \( (v_1) \). The coprimality of \( u_1 \) and \( v_1 \) means \( \mathcal{P} \) \( (s(b))^{2m} \), which gives \( N(\mathcal{P}) \) \( s(b) \), which is impossible because \( s(b) \) is primitive.

Hence, \( \mathcal{P}^{2m} \) \( s(d) \), which means that \( 2m\alpha \leq 2n\beta \). By symmetry, \( 2n\beta \leq 2m\alpha \) and so \( m\alpha = n\beta \). Thus, \( s(b)^m = s(d)^n \) and we find \( N(s(b))^m = N(s(d))^n \). □

Suppose that \( b_1, \ldots, b_n \) are independent ideals of \( \mathcal{O} \) of discriminant \( D \) and let

\[
S = \left\{ \prod_{i=1}^{n} b_i^{a_i} : a_i \geq 0 \text{ for } i = 1, \ldots, n \text{ and } \prod_{i=1}^{n} N(b_i)^{a_i} < \sqrt{D}/2 \right\}
\]

If \( a_i = Q_{i-1}Z + (\omega + P_{i-1})Z \) is a reduced ideal then

\[
\frac{\omega + P_{i-1}}{Q_{i-1}} > \frac{\sqrt{D}/2}{Q_{i-1}} = \frac{\sqrt{D}/2}{|N(a_i)|}.
\]

Now suppose \( (v)a \in S \), where \( a \) is reduced and \( a = (Q_{i-1}, \omega + P_{i-1}) \). We have

\[
\frac{\omega + P_{i-1}}{Q_{i-1}} > \frac{v^2 \sqrt{D}/2}{\prod_{i=1}^{n} N(b_i)^{a_i}} \geq \frac{\sqrt{D}/2}{\prod_{i=1}^{n} N(b_i)^{a_i}}.
\]

**Theorem 2.8.** Let \( \mathcal{O}_1, \mathcal{O}_2, \ldots, \) be a sequence of orders, each of discriminant \( D_i \), where \( D_i < D_{i+1} \). Suppose further that in each \( \mathcal{O}_i \) there exists an independent set of principal ideals \( \{b_{i,j} : (j = 1, \ldots, n)\} \) such that \( N(b_{i,j}) \) is fixed for each value of \( i \). Then

\[
R(D_i) \gg (\log D_i)^{n+1}.
\]

See [Patterson 2003] for a proof.

### 3. Basic observations on keepers

Given discriminants of the form \( D_n = U^2 x^{2n} + V x^n + W^2 \), where \( U, V, W \in \mathbb{Q} \), there is no loss in generality in supposing that \( x \) is not a power. We may write our discriminants \( D_n \) as

\[
(3-1) \quad D_n = \frac{c^2}{d^2} \left( (Ax^n + C)^2 + 4Gx^n \right), \quad \text{where} \quad G = (B - 2AC)/4,
\]

for \( A, B, C, G, c, d \in \mathbb{Z} \) and \((c, d) = 1\).
Any common factors of $C$ and $x$ can be moved into the square divisors. By considering

$$W := \max_{i \in \mathbb{N}} (x^i, C), \quad m := \min \{ i \in \mathbb{N} : (x^i, C) = W \}, \quad v := n - 2m,$$

so that $(x, C/W) = 1$, we have

$$D_n = \frac{c^2W^2}{d^2} \left( \left( AWx^{2m} + \frac{C}{W}\right)^2 + 4Gx^{2m} \right) = \frac{c^2W^2}{d^2} \left( (\overline{A}x^v + \overline{C})^2 + 4\overline{G}x^v \right),$$

where $(x, \overline{C}) = 1$ and $\overline{A}, \overline{C}, \overline{G} \in \mathbb{Z}$. Any square factors of $(A^2, C^2, G)$ can also be removed, so that without loss of generality we may suppose that

$$D_n = \left( \frac{c}{d} \right)^2 ((Ax^n + C)^2 + 4Gx^n)$$

with $A, C, G \in \mathbb{Z}$, $(x, C) = 1$ and $(A^2, C^2, G)$ is squarefree.

The next few results don’t use any of the properties of krees, we are merely interested in determining an explicit formula for $(Ax^n + C)^2 + 4Gx^n$. Consequently, we define

\begin{equation}
E_n := (Ax^n + C)^2 + 4Gx^n,
\end{equation}

where $(x, C) = 1$ and $(A^2, C^2, G)$ is squarefree. The first result is a representation of $A, C, G$.

**Theorem 3.1.** Given $E_n$ as in (3-2) and the conditions on $A, C, G, x$ stated above, we have that

\begin{equation}
E_n = (qrNx^n + P(M - L)/q)^2 + 4rLPx^n,
\end{equation}

where $r, q, P, N \in \mathbb{Z}^+$, $M, L \in \mathbb{Z}$, and the following conditions are satisfied:

1. $r$ is squarefree,
2. $(P, rqN) = 1$,
3. $rq | M - L$,
4. $(M, L) = 1$,
5. $(rq, ML) = 1$,
6. $(r, LPN) = 1$.

**Proof.** The selections we make are

$$r := (A, C, G), \quad N := \left( \frac{A}{r}, \frac{G}{r} \right), \quad q := \frac{A}{rN},$$

$$P := \left( \frac{C}{r}, \frac{G}{rN} \right), \quad L := \frac{G}{rNP}, \quad M := \frac{AC + G}{rNP}.$$

These selections make (3-2) and (3-3) equivalent, so it only remains to show that the conditions indicated hold. This is not difficult; see [Patterson 2003].

Our next objective is to determine the terms that divide $E_n$ and those that are coprime to $E_n$. 

□

92 ROGER D. PATTERSON, ALFRED J. VAN DER POORTEN AND HUGH C. WILLIAMS
Lemma 3.2. \((E_n, xNP) = 1\).

Proof. See [Patterson 2003]. □

Theorem 3.3. For the family of discriminants given by (3-3) with \(n \in I\) (an infinite subset of \(\mathbb{N}\)) there exists an infinite set \(I' \subseteq I\) such that for every \(n \in I'\) we have \(M = v'mz^2Z, L = vy^2Y\) with \(Z, Y, z, y\) positive and

\[
ml \mid E_n', \quad vv'zy \mid F_n, \quad vv' \mid 2, \quad (ZY, E_n') = 1, \quad \left(\frac{F_n}{zy}, mlZY\right) = 1
\]

and \(E_n = F_n^2E_n',\) where \(E_n'\) is squarefree. Here \(m, l, y, z, Y, Y\) are the same for all \(n \in I'\).

Proof. Define \(y_n^2v_n := (F_n^2, L),\) where \(v_n\) is squarefree. Next, write

\[
\tilde{F}_n = \frac{F_n}{y_n v_n} \quad \text{so that} \quad \left(v_n \tilde{F}_n^2, \frac{L}{v_n y_n^2}\right) = 1,
\]

and

\[
E_n = y_n^2 v_n^2 \tilde{F}_n^2 E_n' = S_n^2 + 4r v_n y_n^2 \frac{L}{v_n y_n^2} NP x^n,
\]

where

\[
S_n = qrNx^n + \frac{P(M - L)}{q}.
\]

It is not difficult to show that \(v_n \mid 2;\) see [Patterson 2003].

We now investigate the factors of \(L/v_n y_n^2.\) (Since \((F_n/(v_n y_n), L/(v_n y_n^2)) = 1\) it follows that for any prime factor, \(p,\) of \(L/v_n y_n^2,\) we have \(p \mid S_n/(v_n y_n)\) if and only if \(p \mid E_n').\) We define \(l_n\) to be the product of all prime powers \(p^\alpha\) of \(L/v_n y_n^2\) that satisfy \(p^\alpha \mid L/v_n y_n^2\) and \(l_n E_n'\) and \(Y_n\) to be the product of all prime powers \(p^\beta\) that satisfy \(p^\beta \mid L/v_n y_n^2\) and \((E_n', p) = 1.\) We also absorb the sign of \(L\) into \(l_n.\) Clearly,

\[
L = v_n l_n y_n^2 Y_n, \quad (Y_n, E_n') = 1, \quad v_n y_n \mid F_n, \quad l_n \mid E_n', \quad v_n \mid 2.
\]

Equation (3-4) can also be written as

\[
E_n = z_n^2(v_n')^2 \tilde{F}_n^2 E_n' = (S_n')^2 + 4r v_n' z_n^2 \frac{M}{v_n' z_n^2} NP x^n,
\]

where \(v_n'\) is squarefree and

\[
S_n' = qrNx^n - P(M - L)/q \quad \text{and} \quad z_n^2 v_n' = (F_n^2, M).
\]

By similar reasoning we get \(M = m_n v_n' z_n^2 Z_n, \) \((Z_n, E_n') = 1, \) \(v_n' z_n \mid F_n, \) \(m_n \mid E_n', \)

\(v_n' \mid 2,\) and \(4\mid v_n v_n'\) because \((M, L) = 1.\)
As there are only a finite number of choices for \(m, l, z, y, v, v', Z, Y\) for fixed values of \(L\) and \(M\) there must exist some infinite set \(I' \subseteq I\) for which

\[
M = v'mz^2Z, \quad L = vly^2Y, \quad ml \mid E_n', \quad zyvv' \mid F_n, \quad (ZY, E_n') = 1, \quad vv' \mid 2.
\]

The coprime conditions all follow easily. \(\square\)

This completes our investigation of \(E_n\); we have determined the factors coprime to \(E_n\) and those which divide each part of it.

4. Independent ideals in kreepers

Completing the square on \(A^2E_n\) gives,

\[
A^2E_n = \left(A^2x^n + B/2\right)^2 - 4GH
\]

and \(B/2\) is an integer because \(4 \mid (B - 2AC)\). In terms of the constants found in the previous section, this equation becomes

\[
(4-1) \quad q^2F_n^2E_n' = (q^2rNx^n + P(L + M))^2 - 4P^2LM.
\]

We define \(\omega_n' := (\sigma_n' - 1 + \sqrt{E_n'})/\sigma_n'\), where \(E_n'\) is squarefree and \(\sigma_n' = 2\) if \(E_n' \equiv 1 \pmod{4}\) and \(\sigma_n' = 1\) otherwise. We further define \(C_n = \mathbb{Z} + \omega_n\mathbb{Z}\), where

\[
\omega_n := t_n + \sqrt{D_n}/2 \quad \text{and} \quad t_n = \begin{cases} 0 & \text{if } D_n \equiv 0 \pmod{4}, \\ 1 & \text{if } D_n \equiv 1 \pmod{4}. \end{cases}
\]

Proposition 4.1. \(C_n = \mathbb{Z} + \frac{cF_n\sigma_n'}{2d} \omega_n\mathbb{Z}\).

Our objective now is to find an ideal arising from (4-1) which has a norm coprime to the conductor of some order.

Proposition 4.2. For each \(n\), there exists an element \(\alpha_n \in \mathbb{Z} + f_1\omega_n\mathbb{Z}\), where \(f_1 = F_n/(\lambda z y)\), and \(\lambda = 1\) or 2, and \((N(\alpha_n), f_1) = 1\).

Proof. We take \(\alpha_n\) as

\[
\alpha_n := \begin{cases} s_n/2 - \frac{qF_n}{2zy} + \frac{qF_n}{zy}\omega_n' & \text{if } 2 \nmid q \text{ and } 2 \nmid F_n/zy, \\ s_n/2 + \frac{qF_n}{2zy}\sqrt{E_n'} & \text{otherwise}. \end{cases}
\]

The remaining details are in [Patterson 2003]. \(\square\)
Bounded norms in kreepers. By the definition of a kreeper, the continued fraction expansion of $\omega_n$ has some $Q_i = x^k$ with $g$ fixed independently of $n$. In other words, there exists some $\mu_n \in \mathcal{O}_n$ such that $N(\mu_n) = x^k$. Also recall Proposition 4.1, which states that $\mathcal{O}_n = \mathbb{Z} + f_2 \omega_n\mathbb{Z}$, where $f_2 = c F_n \sigma_n / 2 d$. Let $f_3 = (f_1, f_2)$, so that $\alpha_n$, $\mu_n$ are both contained in $\mathcal{O}_n^* := \mathbb{Z} + f_3 \omega_n\mathbb{Z}$. We know that $(x, E_n(\mu_n), f_3) = (N(\mu_n), f_3) = 1$.

Thus in the order $\mathcal{O}_n^*$ we have the two principal ideals $a_n = (\alpha_n)$ and $b_n = (\mu_n)$ (whose norms are fixed for infinitely many $n \in I$). Both ideals have norms coprime to the conductor. Hence, we may apply Theorems 2.7 and 2.8.

**Proposition 4.3.** If $a_n$ and $b_n$ are independent ideals in $\mathcal{O}_n^*$ then
\[
R(\mathcal{O}_n) \gg (\log D_n)^3.
\]

**Proof.** By Theorem 2.8,
\[
R(\mathcal{O}_n^*) \gg (\log \Delta(\mathcal{O}_n^*))^3 = (\log (f_3^2 \Delta(\mathcal{O}_n^*)))^3,
\]
where $\mathcal{O}_n' := \mathbb{Z} + \omega_n\mathbb{Z}$ and $\Delta(\mathcal{O})$ denotes the discriminant of the order $\mathcal{O}$. By Proposition 2.5,
\[
(4-2) \quad R(\mathcal{O}_n) > \frac{(f_3, f_2)}{2 f_3} R(\mathcal{O}_n^*) \implies R(\mathcal{O}_n) > \frac{1}{2} R(\mathcal{O}_n^*) \gg (\log (f_3^2 \Delta(\mathcal{O}_n^*)))^3.
\]
Since $f_3 = F_n(2d, c \sigma_n \zeta N)/2 \zeta d$ we find
\[
(\log f_3^2 \Delta(\mathcal{O}_n^*))^3 \gg (\log F_n^2 \Delta(\mathcal{O}_n^*))^3 \gg (\log F_n^2 E_n^2)^3 \gg (\log D_n)^3.
\]
Hence, from (4-2), we have that $R(\mathcal{O}_n) \gg (\log D_n)^3$. 

Consequently, in order for the sequence of discriminants $\{D_n\}$ to be a kreeper, $a_n$ and $b_n$ must be dependent ideals in $\mathcal{O}_n^*$. By Theorem 2.7 there must exist nonnegative integers $e$ and $f$, with $e$ and $f$ not both 0 such that $N(s(\alpha_n))^e = N(s(b_n))^f$.

With not too much effort (see [Patterson 2003, Chapter V, §15]) one shows that $N(s(\alpha_n)) = P^2 Z Y$ and $N(s(b_n)) = x^8$. If $e = 0$ then $x^{8f} = 1$ and since in this case we must have $f > 0$ we find $x = 1$, which is impossible. On the other hand, if $f = 0$ then $Y$, $Z$ and $P$ are all $\pm 1$, in which case
\[
\frac{q^2 E_n}{(zy)^2} = s_n^2 \pm 4muv
\]
But $muv'$ divides $E_n/(zy)^2$, which by a result of Schinzel [1961] implies that the period length of $q^2 E_n/(zy)^2$ is bounded for all $n$. In the terminology of [Kaplansky 1998], this says that $q^2 E_n/(zy)^2$ is a sleeper. It is shown in [Patterson
2003, Chapter III] that any rational multiple of a sleeper is again a sleeper. Hence, \( c^2 E_n / d^2 = D_n \) must also be a sleeper; in other words, \( D_n \) is not a kreeper. It follows that if \( D_n \) is to be a kreeper we must have \( N(s(a_n))^f = N(s(b_n))^g \) with \( e, f \) each being positive. Since \( (x, P) = 1 \) we must have \( P = \pm 1 \). Hence, we may replace \( PM \) by \( M \) and \( PL \) by \( L \) and all the previous conditions hold. Taking \( d' := (gf, e), k := gf / d, h := e / d \) gives,

\[
(ZY)^b = x^k, \quad (h, k) = 1,
\]

which implies that \( x = R^h, ZY = R^k \). Hence, \( h = 1 \) and \( ZY = x^k \) because \( x \) is not a power. From Theorem 3.1 we have \( (Z, Y) = 1 \); hence

\[
Z = U^k, \quad Y = T^k, \quad \text{where } (U, T) = 1.
\]

The objective of the next 2 sections will be to show that \( T = 1 \) and \( U = x \).

5. Part of the continued fraction expansion of \( \omega_n \)

There is no longer any need to distinguish between factors of \( E'_n \) and \( F_n \), which means that we may absorb the terms \( v \) and \( v' \) into \( l \) and \( m \) respectively. The form of a kreeper is now given by

\[
D_n = \left( \frac{c}{d} \right)^2 \left( (qr N(UT)^a + (z^2 m U^k - y^2 l T^k) / q)^2 + 4r N l y^2 T^{k+a} U^n \right),
\]

where \( m, l, r \) are squarefree and

\[
(qr, UT) = 1, \quad (qr NUT, yzm) = 1, \quad (Tyl, mzU) = 1, \quad qr \mid z^2 m U^k - y^2 l T^k
\]

and for every \( n \in I \),

\[
(yl \mid q^2 r N x^a + z^2 m U^k \quad \text{and} \quad mz \mid q^2 r N x^a + y^2 l T^k)
\]

and \( N > 0, x = UT \). Let \( \mu \) be the least positive difference of any two integers in \( I \). Then \( v, v + \mu \in I \) for some \( v \), and

\[
yl \mid q^2 r N x^v + z^2 m U^k \quad \text{and} \quad yl \mid q^2 r N (U^v x^{v+\mu} + z^2 m U^k),
\]

which means \( yl \mid q^2 r N x^v (x^{\mu} - 1) \), so \( yl \mid x^{\mu} - 1 \) because \( (yl, qr N x) = 1 \). By symmetry, \( zm \mid x^{\mu} - 1 \). Thus, \( ylmz \mid x^{\mu} - 1 \). Hence the conditions (5-2) become

\[
yl \mid q^2 r N (U^v)^{x^\mu} + z^2 m U^k, \quad mz \mid q^2 r N (U^v) + y^2 l T^k \quad \text{and} \quad (TU)^{x^\mu} \equiv 1 \pmod{ylmz}
\]

for any \( n \in I \), such that \( n \equiv v \pmod{\mu} \). Since the signs of \( m \) and \( l \) have not yet been specified, there is no loss of generality in supposing that \( q, U, T \) are all positive.
Further, we have where we may assume without loss of generality that $U > T$.

**Some notation.** We now rewrite equations (5-1) and (5-3) as

\[ D_n = \left( \frac{s_1}{d} \right)^2 + 4 \left( \frac{c}{d} \right)^2 r N \gamma^2 T^{k+n} U^n = \left( \frac{s_2}{d} \right)^2 + 4 \left( \frac{c}{d} \right)^2 r N m z^2 T^{k+n} U^{k+n}, \]

where

\[ s_1 = c q r N x^n + c(z^2 m U^k - y^2 I T^k)/q, \quad s_2 = c q r N x^n - c(z^2 m U^k - y^2 I T^k)/q. \]

We choose an infinite subset of $I$ such that $t \equiv D_n \pmod{4}$ is fixed. Then we take $S_1 := (s_1 - td)/2$ and $S_2 := (s_2 - td)/2$. We also write

\[ \alpha = \omega_n + S_1/d, \quad \alpha = \alpha_n + S_1/d \quad \text{and} \quad \beta = \omega_n + S_2/d, \quad \beta = \alpha_n + S_2/d. \]

Then

\[ \alpha \alpha = -c^2 r N \gamma^2 T^k x^n / d^2 \quad \text{and} \quad \beta \beta = -c^2 r N m z^2 T^k x^n / d^2. \]

Further, we have

\[ q^2 d^2 D_n = (qs_3)^2 - 4 c^2 ml (zy)^2 (UT)^k, \]

where

\[ qs_3 = c q^2 r N x^n + m z^2 U^k + l y^2 T^k. \]

Also of relevance will be $S_1 + S_2 + td = Ax^n = c q r N x^n$.

We now detail an initial segment of the continued fraction expansion of $\omega_n$. In the case of $T > 1$, this segment will have length $O(n^{1 + \varepsilon})$, hence the entire expansion could not satisfy $lp(\omega_n) = an + b$, as required by kקלאers.

Before commencing we need to determine the common factors between some of the terms. First, we define $g := (s_1, s_2, d)$. It is not to difficult to show that $(z, s_1/c)$ and $(y, s_2/c)$ each divides 2. We also define $d_y := (S_1, d)$, $d'_y := (s_1/g, d/g)$ and $\tau_y := d_y/d'_y$. It follows easily that $\tau_y$ is an integer. Similarly, we define $d_z := (S_2, d)$, $d'_z := (s_2/g, d/g)$ and $\tau_z := d_z/d'_z$. Next, we write $d = \tilde{d} d_x d_y$. Here are some simple results:

- $g \mid 2$, moreover, $g = 2$ if and only if $2 \mid d$.
- $\tau_c \mid g$ and $\tau_y \mid g$.
- $\tau_y d'_z \mid y$ and $\tau_z d'_y \mid z$; in other words $d_y \mid y$ and $d_z \mid z$. 
Expansion of $\omega_n$. The continued fraction expansion of $\omega_n$ begins as

$$\omega_n = \frac{(S_1 + td)/d_y}{d/d_y} - \left(\frac{\omega_n + S_1}{d}ight)$$

and $((S_1 + td)/d_y, d/d_y) = 1$ by the definition of $d_y$. Hence we may apply Lemma 2.2, and find that after the expansion of $(S_1 + td)/d$, of length $h_0$, a new complete quotient in the continued fraction expansion is

$$-(-1)^{h_0+1} \frac{c_0}{(\omega_n + S_1/d)(d/d_y)^2} - \frac{c_0}{d/d_y},$$

where

$$c_0 \equiv \frac{(-1)^{h_0}}{(S_1 + td)/d_y} \pmod{d/d_y}.$$

By choosing $(-1)^{h_0+1} = \text{sign}(l)$ the element in (5-4) then becomes

$$\frac{\omega_n + S_1/d}{c_2 r N |l| y^2 T^k x^n / (d/y)^2} - \frac{c_0}{c_2 r N |l| (y/d_y)^2 T^k x^n} = \frac{\omega_n + S_1/d - c_0 c_2 r N |l| d_y (y/d_y)^2 T^k x^n / d}{c_2 r N |l| (y/d_y)^2 T^k x^n}.$$

Now define: $s := \max_{i \in \mathbb{N}} \{ (x^n, c) \}$ and $u := c/s$. Hence $(u, x) = 1$. Recall, $(x, qrzmyl) = 1$ so then $(s, qrzmyl) = 1$. We will denote the element in (5-5) as $\theta_{h_0}$. From it, we can write

$$\theta_{h_0} = \frac{\omega_n + P_{h_0}}{Q_{h_0}} = \frac{A_0}{B_0} = \frac{\beta}{c_2 r N |l| (y/d_y)^2 T^k x^n},$$

where

$$A_0 = c q r N x^n - c_0 c_2 r N |l| d_y (y/d_y)^2 T^k x^n, \quad B_0 = c_2 r N |l| (y/d_y)^2 T^k x^n.$$

Next, we define $\Delta_0 := (A_0, B_0)$. We need to determine $\Delta_0$ before we can apply Lemma 2.2 again. One finds,

$$\Delta_0 = c r N x^n \delta d_z \quad \text{and} \quad B_0/\Delta_0 = c/\delta d/d_z |l| (y/d_y)^2 T^k.$$

From (5-6), by applying Lemma 2.2, we find the next partial quotients are those of the continued fraction expansion of $A_0/B_0$ of length $p_0$. By choosing $(-1)^{p_0+1} = \text{sign}(m)$, the next element in the continued fraction expansion is $\theta_{h_1}$, where $h_1 := h_0 + p_0$,

$$\theta_{h_1} = \frac{c_2 r N |l| (y/d_y)^2 T^k x^n}{-\beta \text{sign}(m) (B_0/\Delta_0)^2} - \frac{c_1}{B_0/\Delta_0}.$$
and $c_1 \equiv -\text{sign}(m) \Delta_0/A_0 \pmod {B_0/\Delta_0}$. We can write

\begin{equation}
\theta_h = \frac{\omega_n + S_2/d - c_1(c/\delta)|m|z(dz/dc)U^k/d}{(c/\delta)^2|m|(y/dy)^2(z/dz)^2(UT)^k}.
\end{equation}

**Finding other complete quotients.** The set of conditions

$$u_{2i-1} \mid u, \quad z_{2i-1} \mid z/dz, \quad z_{2i-1}' \mid z, \quad (c_{2i-1}, sT) = 1, \quad (z_{2i-1}y_{2i-1}, u/u_{2i-1}) = 1,$$

$$m_{2i-1} \in \{1, |m|\}, \quad l_{2i-1} \in \{1, |l|\}, \quad r_{2i-1} \in \{1, r\}, \quad y_{2i-1} \mid y/dy, \quad u_{2i-1}' \mid u.$$

will be denoted by $C_{2i-1}$. The set of conditions $C_{2i}$ are the same as $C_{2j-1}$ (with $2j - 1$ replaced by $2i$) except that instead of requiring $z_{2j-1}' \mid z$, we need $y_{2i}' \mid y$.

**Theorem 5.1.** Suppose that there exists a complete quotient $(\omega_n + P_{h_{2i-1}})/Q_{h_{2i-1}}$ satisfying

\begin{align}
(5-8a) \quad P_{h_{2i-1}} &= S_2/d - sm_{2i-1}r_{2i-1}u_{2i-1}u_{2i-1}'z_{2i-1}c_{2i-1}U^k/d \\
(5-8b) \quad Q_{h_{2i-1}} &= r_{2i-1}l_{2i-1}m_{2i-1}(su_{2i-1}y_{2i-1}z_{2i-1})^2(UT)^k,
\end{align}

where $n > ki$ and the set of conditions $C_{2i-1}$ are satisfied. Then there is a complete quotient $(\omega_n + P_{h_2})/Q_{h_2}$, where

\begin{align}
(5-9a) \quad P_{h_2} &= S_1/d - Nr_{2i}l_{2i}u_{2i}u_{2i}'y_{2i}c_{2i}U^{n-ki}T^{n+k}/d \\
(5-9b) \quad Q_{h_2} &= r_{2i}l_{2i}m_{2i}(su_{2i}y_{2i}z_{2i})^2NT^{n+k(i+1)}U^{n-ki}
\end{align}

and the conditions $C_{2i}$ are satisfied.

Observe that with appropriate selections, $\theta_{h_1}$ is a complete quotient satisfying (5-8) and the conditions $C_1$.

**Proof.** From (5-8), we find that the current line in the continued fraction expansion is

$$\frac{\omega_n + P_{h_{2i-1}}}{Q_{h_{2i-1}}} = \frac{cqrN x^n - sr_{2i-1}m_{2i-1}u_{2i-1}u_{2i-1}'z_{2i-1}c_{2i-1}U^k}{dr_{2i-1}m_{2i-1}l_{2i-1}(su_{2i-1}z_{2i-1}y_{2i-1})^2(UT)^k} \cdot \frac{(\omega_n + S_1/d)}{r_{2i-1}m_{2i-1}l_{2i-1}(su_{2i-1}z_{2i-1}y_{2i-1})^2(UT)^k}.$$

Now, define

$$A_{2i-1} := cqrN x^n - sr_{2i-1}m_{2i-1}u_{2i-1}u_{2i-1}'z_{2i-1}c_{2i-1}U^k,$$

$$B_{2i-1} := dr_{2i-1}m_{2i-1}l_{2i-1}(su_{2i-1}z_{2i-1}y_{2i-1})^2(UT)^k,$$

$$\Delta_{2i-1} := (A_{2i-1}, B_{2i-1}).$$

The next few results aid in determining common factors.

**Lemma 5.2.** If $A_{2i-1} = dP_{h_{2i-1}} + (s_1 + td)/2$ then $d_yu_{2i-1}l_{2i-1}y_{2i-1} \mid A_{2i-1}$.
Lemma 5.3. For any rational integers $a$, $b$, $d$, $f$ such that $d \mid ab$ and $(f, d) = 1$ there exist rational integers $x$, $y$ such that

$$dxy = ab$$

and $x \mid a$, $y \mid b$, $(fx, b/y) = 1$.

Proof. Take $x = a/(a, d)$ and $y = b(a, d)/d$. \qed

Returning to the expansion of $\omega_n$, we write

$$w_{2i-1} := (m_{2i-1}, u/2_{2i-1}) \quad \text{and} \quad e_{2i-1} := s_{i}\frac{u_{2i-1}}{2_{2i-1}}y_{2i-1}r_{2i-1}w_{2i-1}.$$ 

Note that if $n \geq ki$ then $w_{2i-1}u_{2i-1}sT^i r_{2i-1} \mid A_{2i-1}$ and by Lemma 5.2 we get $d_{i}u_{2i-1}l_{2i-1}y_{2i-1}z_{2i-1} \mid A_{2i-1}$. Now we define $g_{2i-1} := A_{2i-1}/(U^i e_{2i-1})$. It follows easily (see [Patterson 2003, Chapter 16]) that $(g_{2i-1}, sT_{z_{2i-1}}m_{2i-1}/w_{2i-1}) = 1$.

In summary,

$$\Delta_{2i-1} = U^i e_{2i-1}(G_{2i-1}, \bar{d}_{2i}y_{2i-1}u_{2i-1}).$$

Since $A_{2i-1} = dP_{h_{2i-1}} + S_{1} + td$, we have $(d, A_{2i-1}) = d_{1}$. Thus $(\bar{d}_{2i}, G_{2i-1}) = 1$.

Hence, $\Delta_{2i-1} = U^i e_{2i-1}\Delta_{2i-1}$, where $\Delta_{2i-1} := (y_{2i-1}u_{2i-1}, G_{2i-1})$.

From the complete quotient

$$\theta_{h_{2i-1}} = \frac{A_{2i-1}}{B_{2i-1}} - \frac{\bar{a}}{r_{2i-1}l_{2i-1}m_{2i-1}(su_{2i-1}y_{2i-1}z_{2i-1})^2(U^i T^i)}$$

we apply Lemma 2.2, so the next partial quotients are those of the expansion of $A_{2i-1}/B_{2i-1}$ of length $p_{2i-1}$. The parity of $p_{2i-1}$ is determined by $(-1)^{p_{2i-1}+1} = \text{sign}(d_{1})$. Following this, the next complete quotient is $\theta_{h_{2i}}$, where $h_{2i} := h_{2i-1} + p_{2i-1}$. By Lemma 2.2, $\theta_{h_{2i}}$ is equal to

$$\frac{r_{2i-1}l_{2i-1}m_{2i-1}(su_{2i-1}y_{2i-1}z_{2i-1})^2(U^i T^i)}{\bar{a}(B_{2i-1}/\Delta_{2i-1})^2} - \frac{c_{2i}}{B_{2i-1}/\Delta_{2i-1}} = \frac{r_{2i-1}l_{2i-1}(a_{n} + S_{1}/d_{1})(w_{2i-1}d_{2i}z_{2i-1})^2}{s^{2}u^{2}rN(|y^{2}m_{2i-1}z_{2i-1}^{2}U^{n-k_{1}}T^{n-k_{1}+1})} - \frac{c_{2i}}{B_{2i-1}/\Delta_{2i-1}},$$

where

$$c_{2i} \equiv - \text{sign}(l_{1})\Delta_{2i-1}/A_{2i-1} \pmod{B_{2i-1}/\Delta_{2i-1}}.$$ 

Also note that $sT_{z_{2i-1}}m_{2i-1}/w_{2i-1}$ and $(A_{2i-1}/\Delta_{2i-1}, B_{2i-1}/\Delta_{2i-1}) = 1$ imply that $(c_{2i}, sT) = 1$.

According to Lemma 5.3 there exists $y_{2i}$, $u_{2i}$ such that

$$y_{2i} \mid y/d_{1}, \quad u_{2i} \mid u/w_{2i-1}, \quad y_{2i}u_{2i}\Delta_{2i-1} = \frac{y}{d_{1}}w_{2i-1}$$

and $(z_{2i-1}y_{2i}, u/(u_{2i}w_{2i-1})) = 1$. By taking $l_{2i} := |l|/l_{2i-1}$, $r_{2i} := r/r_{2i-1}$, $m_{2i} := m_{2i-1}$, $z_{2i} := z_{2i-1}$, $y_{2i}' := y/y_{2i-1}$, and $u_{2i}' := u/u_{2i-1}$, one finds that
Suppose there is a complete quotient \( U \). Moreover, the complete quotient \( \theta_{h_2} \) now satisfies (5-9) and the set of conditions \( C_{2i} \).

There is also an analogous result for \( \theta_{h_2} \).

**Theorem 5.4.** Suppose there is a complete quotient \( \theta_{h_2} = (\omega_n + P_{h_2})/Q_{h_2} \) satisfying (5-9) and the set of conditions \( C_{2i} \). Then there is a complete quotient \( \theta_{h_{2+1}} \), where

\[
\begin{align*}
P_{h_{2+1}} &= S_2/d - sm_{2i+1}r_{2i+1}u_{2i+1}u_{2i+1}c_{2i+1}U^{k(i+1)}/d \\
Q_{h_{2+1}} &= r_{2i+1}m_{2i+1}(su_{2i+1}c_{2i+1}T^{n+k}(UT)^{k(i+1)})^2
\end{align*}
\]

and the set of conditions \( C_{2i+1} \) are satisfied.

Thus, from the complete quotient \( \theta_{h_{2-1}} \), we find another complete quotient \( \theta_{h_{2+1}} \) satisfying exactly the same requirements as \( \theta_{h_{2-1}} \). Moreover, only a bounded number of these complete quotients are not reduced. More precisely,

\[
\frac{A_{2i}}{B_{2i}} = \frac{k_1x^n - k_2U^{n-ki}T^{n+k}}{k_3T^{n+k(i+1)}U^{n-ki}} = k_1U^{ki}k_3T^{(k(i+1))} > 1
\]

for some \( i \geq W \), where \( W \) only depends on the parameters \( m, l, y, z, r, d, c, N, U, T \).

Similarly,

\[
\frac{A_{2i+1}}{B_{2i+1}} = \frac{k_1x^n - k_2U^{k(i+1)}}{k_3(UT)^{k(i+1)}} = \frac{k_1x^n - k_2U^{k}}{k_3(UT)^{k}} = \frac{k_1x^{n-ki}}{k_3T^{k}} > 1
\]

for some \( V \) such that \( n - ki \geq V \). Again, \( V \) depends only on the parameters \( m, l, y, z, r, d, c, N, U, T \).

Since the pairs \( (P_{h_i}, Q_{h_i}) \) are all distinct for \( i = 1, 2, \ldots, 2(n - V)/k + 1 \), we see that

\[
lp(\omega_n) > \sum_{j=W}^{(n-V)/k} p_{2j}.
\]

Our interest now falls on the length of the expansion of \( A_{2i}/B_{2i} \). Basically, since we have \( \Theta(n) \) of these expansions, if the lengths are unbounded then, from above, the period length can not possibly be linear in \( n \). In the next section we show that in order to have the length of \( A_{2i}/B_{2i} \) bounded for all \( i \) we must have \( T = 1 \).

### 6. The length of the continued fraction of \( A_{2i}/B_{2i} \)

Let \( i \) be fixed with \( W \leq i \leq (n - V)/k \), so that \( \theta_{h_2} \) is reduced. We have

\[
P_{h_2} = S_1/d - Ns_{2i}r_{2i}u_{2i}u_{2i}c_{2i}U^{n-ki}T^{n+k}/d = S_1/d - J/d.
\]
As usual, we have
\[ d^2(t - D_n)/4 + d^2(t P_{h_2i} + P_{h_2i}^2) \equiv 0 \pmod{d^2 Q_{h_2i}}, \]
where in the above case
\[ Q_{h_2i} = r_{2i} m_{2i} l_{2i} N (su_{2i} z_{2i} y_{2i})^2 U^{n - ki} T^{n+k(i+1)}. \]
We can write this as
\[ -Gx^n - td J - 2S_1 J + J^2 \equiv 0 \pmod{d^2 Q_{h_2i}}. \]
Modulo \( T^{k(i+1)} \) we have
\[(6-1) \quad GU^{ki} \equiv -CNsl_{2i} r_{2i} u_{2i} y_{2i} c_{2i} T^k \pmod{T^{k(i+1)}}.\]
Returning now to \( A_{2i}/B_{2i} \), we previously found that
\[ \frac{A_{2i}}{B_{2i}} = \frac{AU^{ki} - Nr_{2i} l_{2i} u_{2i} y_{2i} c_{2i} T^k}{dN l_{2i} r_{2i} m_{2i} (su_{2i} z_{2i} y_{2i})^2 T^{k(i+1)}}. \]
Writing (6-1) as
\[ C Nsl_{2i} r_{2i} u_{2i} y_{2i} c_{2i} T^k = -GU^{ki} - f_{2i} T^{k(i+1)}, \]
where \( f_{2i} \in \mathbb{Z} \), we find,
\[ \frac{A_{2i}}{B_{2i}} = \frac{c^2 r N m z^2 U^{k(i+1)} + f_{2i} T^{k(i+1)}}{C N d l_{2i} r_{2i} m_{2i} (s u_{2i} z_{2i} y_{2i})^2 T^{k(i+1)}} = \frac{1}{F_{2i}} \left( E \xi^{i+1} + f_{2i} \right), \]
where \( F_{2i} \) and \( E \) are bounded integers and \( \xi = U^k / T^k \).

**Depth of a sequence of rationals.** We now provide an aside regarding the depth of \( (a/b)^h \) as \( h \to \infty \) for coprime integers \( a \) and \( b \). The depth of the regular continued fraction expansion of \( \alpha \in \mathbb{Q} \) is denoted by \( \delta(\alpha) \). This is defined as the number of partial quotients in the even length continued fraction expansion of \( \alpha \).

**Theorem 6.1.** If \( a \) and \( b \) are two coprime integers with \( 1 < b < a \) then
\[ \lim_{i \to \infty} \delta \left( (a/b)^i \right) = \infty \]
Whether this is so was asked by Mendes France and proved by Pouchret in a letter to him. A summary of Pouchret’s response is given in [van der Poorten 1984].

---

3Van der Poorten provides the following correction to the given argument: Consider \( p_{n+1} < p_n \alpha^{h_{n+1}} \), so that \( a^h = p_{\psi(h)} < a^{h(\xi_1 + \cdots + \psi(h))} \), and then consider \( \xi_1 + \cdots + \psi(h) = \psi(h) \).
It is clear that $\delta(1/\alpha) \geq \delta(\alpha) - 2$, $\delta(-\alpha) \geq \delta(\alpha) - 2$ and $\delta(\alpha + n) = \delta(\alpha)$ for any $n \in \mathbb{Z}$. Furthermore, Mendès France [1973] has shown that for any $\alpha \in \mathbb{Q}, n \in \mathbb{Z}^+$,

$$\delta(n\alpha) \geq \frac{\delta(\alpha) - 1}{\kappa(n) + 2} - 1,$$

where $\kappa(n)$ is a positive valued function whose values depend only on $n$. Consequently, for any sequence $\alpha_i \in \mathbb{Q}$ satisfying $\lim_{i \to \infty} \delta(\alpha_i) = \infty$ and any $n \in \mathbb{Z}^+$ we have $\lim_{i \to \infty} \delta(\alpha_i/n) = \infty$. It follows that, if $T > 1$, then

$$\lim_{i \to \infty} \delta \left( \frac{A_{2i}}{B_{2i}} \right) = \lim_{i \to \infty} \delta \left( \frac{1}{F_{2i}} (E_2^{i+1} + f_{2i}) \right) = \infty. \tag{6-2}$$

**Period length of $\omega_n$.** By our criteria for a kreeper we must have for some $a, b \in \mathbb{Q}$,

$$an + b = lp(\omega_n) > \sum_{j=W}^{(n-V)/k} p_{2j}$$

for all $n \in I$. Hence we must have

$$\sum_{i=W}^{(n-V)/k} \delta \left( \frac{A_{2i}}{B_{2i}} \right) < an + b. \tag{6-3}$$

By (6-2) there exists a $\gamma > W$ such that $\delta \left( \frac{A_{2i}}{B_{2i}} \right) > k(a + 1)$ for all $i \geq \gamma$. Then,

$$\sum_{i=W}^{(n-V)/k} \delta \left( \frac{A_{2i}}{B_{2i}} \right) > k(a + 1)[(n - V)/k - \gamma]. \tag{6-4}$$

When $n > b + (a + 1)(V + k\gamma)$ we have (6-4) is greater than $an + b$. And since all the terms on the right side of this inequality are bounded, there must exist an infinitude of values of $n \in I$ such that $lp(\omega_n) > an + b$ for any fixed $a, b$. In conclusion $D_n$ can not be a kreeper if $T > 1$. In other words, we must have $T = 1$ and $U = x$.

Our only remaining objective is to show that necessarily $N = 1$. Clearly, there is no loss in generality in supposing that $N$ is not a power of $x$. In Section 5 we established the existence of the following complete quotient in all kreepers,

$$\theta_{n/2^i} = \frac{\omega_n + S_1/d - Ns_r2_{l2}u_{2i}u_{2i}y_{2i}z_{2i}x^{n-ki}/d}{r_{2i}m_{2i}(su_{2i}z_{2i})^2N \rho x^{n-ki}}.$$  

By taking $i = \lfloor n/k \rfloor$ we find an element, $\eta \in \mathfrak{O}_n$, whose norm can be written as $RN x^\gamma$, where $R \left| D_n \right| (N, E_n) = 1$ and $0 \leq \gamma < k$. Hence the norm of $\eta$ is bounded independently of $n$, and coprime to the conductor of the order $\mathfrak{O}_n^*$. First, recall the ideal $b_n \in \mathfrak{O}_n^*$ from page 195, which has norm $x^g$, with $g$ fixed independently of
n. Suppose that \( N > 1 \) and that the ideals \( (\eta) \), \( b_n \) are dependent in \( \mathcal{O}_n^* \). Then by Theorem 2.7 we must have \( N^e x^e = x^{ \frac{g}{f} } \) for some nonnegative integers \( e, f \), not both zero. Since \( N \) is not a power of \( x \), we must have \( e = 0, f > 0 \). But this would imply \( x = 1 \), which is impossible. Hence \( (\eta) \) and \( b_n \) are independent ideals in \( \mathcal{O}_n^* \).

Then by Proposition 4.3, \( R(\mathcal{O}_n) \gg (\log D_n)^3 \); that is the sequence of discriminants \( \{D_n\} \) is not a kreeper. Thus we must have \( N = 1 \).

This completes the proof of Theorem 1.1.

### 7. Function field kreepers

Almost everything done here is also valid in function fields. Instead of considering quadratic fields over \( \mathbb{Q} \), we can consider the so-called congruence quadratic function fields \( \mathbb{F}_q(X, \sqrt{f}) \), where \( f \in \mathbb{F}_q[X] \). It is well-known that the results in Section 2 have similar analogues for expansions over \( \mathbb{F}_q[X] \). The main exception for our interests here, is that the continued fraction expansion of a rational function \( f(X) \in \mathbb{F}_q[X] \) has a fixed length. This was used in Section 5 to ensure each \( Q_{b_j} \) was positive. In the function field case the term \((-1)^{p_{2n-1}+1} l\) is equal to \( ul \), where \( u \in \mathbb{F}_q^* \). By interpreting \(|l|\) to be equal to \( ul \) for some \( u \in \mathbb{F}_q^* \), the results carry over.

Other minor details: \( P \) lies in \( \mathbb{F}_q^* \) rather than in \( \{\pm 1\} \), which is easily handled by just renaming \( M \); further, \( L, g \) should be defined to be 2 and \( \tau_y \), \( \tau_z \) can be ignored.

The main problem is that Theorem 6.1 does not hold for coprime functions in \( \mathbb{F}_q[X] \). Consequently, we have no direct proof that \( T = 1 \) and \( U = x \). However, if we suppose only the weaker condition: that \( x \) is a monomial in \( X \) (then necessarily \( x = X \) by assumption about powers of \( x \)) then trivially, \( T \in \mathbb{F}_q \) and \( U = x/T \). Then by renaming \( l \) and \( m \) we have:

**Theorem 7.1.** A function field kreeper, that is a sequence of polynomials \( f_n(X) \) such that

\[
(1) \quad f_n(X) = A(X)^2 X^{2n} + B(X) X^n + C^2, \quad \text{where} \quad A, B, C \in \mathbb{F}_q[X]
\]

\[
(2) \quad lp(\sqrt{f_n(X)}) = an + b \quad \text{for some} \quad a, b \in \mathbb{Q}
\]

\[
(3) \quad \text{In the principal cycle there exists an element whose norm is } X^g \text{ for some } g \text{ fixed independently of } n.
\]

must satisfy

\[
d^2 f_n(X) = c^2 \left( (qr X^n + (mz^2 X^k - 1y^2)/q)^2 + 4ry^2 X^n \right),
\]

where \( q, r, l, m \in \mathbb{F}_q[X] \) and

\[
(qr X, mlzy) = 1, \quad (ml, zy) = 1, \quad (qr, X) = 1, \quad c^2 rly^2 mz^2 \big| d^2 D_n, \quad q \big| mz^2 X^k - 1y^2.
\]
8. Some more notations

Define \( s := \max_{i \in \mathbb{N}} \left\{ \langle x^i, c \rangle \right\} \), \( u := c/s \) so that \( (u, x) = 1 \). In order to make things easier for ourselves when we come to the expansion of \( \omega_n \), we would like to have

\[
x^u \equiv 1 \pmod{u^2 mz^2 l^2 y^2} \quad \text{and} \quad x^u \equiv 0 \pmod{s^2}.
\]

Clearly, such a \( \mu \) must exist. We shall want to consider the congruence class, \( I_\nu := \{ n \in \mathbb{N} : n \equiv \nu \pmod{\mu}, n > \mu \} \). Our proof will show that \( D_n \) is a kreeper for \( n \in I_\nu \), and since every \( n \) lies in some \( I_\nu \), we shall not be losing any generality in this restriction. Moreover, the value of \( x^n \pmod{u^2 s^2 mz^2 l^2 y^2} \) is the same for all \( n \in I_\nu \).

If \( \theta_{i+1} \) represents the \((i + 1)\)-th complete quotient of \( \omega_n \), that is, if

\[
\omega_n = [a_0, a_1, \ldots, a_i, \theta_{i+1}]
\]

we define

\[
\Psi_{i+1} = \theta_1 \cdots \theta_{i-1} \theta_i \ B_i \in \mathbb{Z}[\omega_n].
\]

Then, we have \( N(\Psi_{i+1}) = (-1)^i Q_i \). If we write the complete quotient \( \theta_{hi} = (\omega_n + P_{hi})/Q_{hi} \) as \( \theta_{hi} = A_i/B_i - \gamma \), where we take

\[
B_i = d Q_{hi} \quad \text{and} \quad A_i = \begin{cases} d P_{hi} + c q r x^n - S_m & \varepsilon_i = 0 \pmod{2}, \\ d P_{hi} + s_1 - S_2 & \varepsilon_{i-1} = 1 \text{ and } i \equiv 1 \pmod{2}, \end{cases}
\]

where \( P_{hi} = S_m/d - J, S_m \) being one of \( S_1 \) and \( S_2 \) and \( J \) is some function of \( r, l, m, z, y, c, x \). Applying Lemma 2.2 (where \( \Delta_i = (A_i, B_i) \)), we find

\[
\theta_{hi+1} = \frac{(-1)^h \sqrt{\gamma}}{\sqrt{\gamma} (B_i/\Delta_i)^2} - \frac{c_i}{B_i/\Delta_i},
\]

where \( c_i \equiv (-1)^{\rho_i+1} \Delta_i/A_i \pmod{B_i/\Delta_i} \). Thus,

\[
Q_{hi+1} = (-1)^{\rho_i} \sqrt{\gamma} (B_i/\Delta_i)^2 Q_{hi}.
\]

Hence,

\[
\frac{(-1)^{\rho_i} Q_{hi+1}}{Q_{hi}} = \frac{(A_i - B_i \theta_{hi}) (A_i - B_i \bar{\theta}_{hi})}{\Delta_i^2}.
\]

By Corollary 2.4,

\[
\theta_{hi+1} \cdots \theta_{hi+2} = \frac{(A_i - \bar{\theta}_{hi} B_i) Q_{hi}}{\Delta_i Q_{hi+1}}.
\]

Hence,

\[
(8-1) \quad \Psi_{hi+1} = \left( \frac{A_i - B_i \bar{\theta}_{hi}}{\Delta_i} \right) \Psi_{hi+1}.
\]
9. Determination of some specific elements in the expansion

In Section 5 we determined the existence of the following complete quotient in the expansion of \( \omega_n \),

\[
\theta_{h_1} = \frac{\omega_n + S_2/d - C_1(c/\delta)z(z/d_x)x^k/d}{(c/\delta)^2[ml/2](y/d_i)^2(z/d_z)^2x^k},
\]

Moreover, for sufficiently large \( n \), \( \theta_{h_1} \) is reduced. Furthermore,

\[
\Psi_{h_1+1} = \alpha \beta / (d \Delta_0).
\]

As shown earlier, the development of the expansion of \( \omega_n \) depends upon whether or not a power of \( x^k \) can be factored out from \( Q_{h_i} \) or not. In order to accommodate this we define

\[
\varepsilon_i = \begin{cases} 1 & \text{if } \lambda_i \geq n - k, \\ 0 & \text{if } \lambda_i < n - k, \end{cases}
\]

where \( \lambda_i \) is defined recursively as

\[
\lambda_1 := k, \quad \lambda_{i+2} := \begin{cases} \lambda_i + k - n \varepsilon_i & \text{if } i \equiv 1 \pmod{2}, \\ 0 & \text{if } i \equiv 0 \pmod{2}. \end{cases}
\]

Note that \( \lambda_{2i-1} \equiv ki \pmod{n} \).

**Theorem 9.1.** Suppose there exists a complete quotient \( (\omega_n + P_{h_i})/Q_{h_i} \) satisfying

\[
P_{h_i} = S_2/d - sm_i r_i u_i u_i' z_i z_i' c_i x^{\lambda_i}/d, \quad Q_{h_i} = r_i l_i m_i (s u_i y_i z_i)^2 x^{\lambda_i},
\]

and the set of conditions \( C_i \) are satisfied. Then there exists a complete quotient \( (\omega_n + P_{h_{i+2}})/Q_{h_{i+2}} \) satisfying (9-3) and \( C_{i+2} \). Furthermore,

\[
\Psi_{h_{i+2}+1} = \left( \frac{\alpha^{\varepsilon_i+1} \beta}{d^2 \Delta_i \Delta_{i+1}} \right) \Psi_{h_{i+1}}
\]

Note: It is clear that with the appropriate selections, the complete quotient \( \theta_{h_1} \) in (9-1) satisfies the conditions of the theorem.

**Proof.** In Section 5 we determined the existence of a complete quotient satisfying the conditions \( C_{i+1} \). Furthermore,

\[
\Psi_{h_{i+1}+1} = \left( \frac{A_i - B_i \theta_{h_i}}{\Delta_i} \right) \Psi_{h_{i+1}} = \left( \frac{\alpha}{d \Delta_i} \right) \Psi_{h_{i+1}}.
\]

Moreover we have \( h_{i+1} = h_i + p_i \), where \( p_i \) is the length of the appropriately selected continued fraction expansion of \( A_i/B_i \).

Suppose \( \varepsilon_i = 0 \). If \( \lambda_i < n - k \) then \( \varepsilon_i = 0 \) and \( \lambda_{i+2} = \lambda_i + k \). This now follows as in Section 5.
Suppose $\varepsilon_i = 1$. When $\lambda_i \geq n - k$ we have $\varepsilon_i = 1$ and this situation has not yet been considered. In this situation, the previous choice of $A_i+1$ is not appropriate. Instead, we need to consider

$$\theta_{h_i+1} = \frac{s_I/d - sr_i+1l_{i+1}u_{i+1}u'_{i+1}y_{i+1}y'_{i+1}c_{i+1}x^{n-\lambda_i}/d}{r_i+1l_{i+1}m_{i+1}(su_{i+1}y_{i+1}z_{i+1})^2x^{n-\lambda_i}} - \frac{\bar{\alpha}}{r_i+1l_{i+1}m_{i+1}(su_{i+1}y_{i+1}z_{i+1})^2x^{n-\lambda_i}}.$$

There is a slight problem in notation because there is going to be an extra intermediate complete quotient. Consequently, we will use overlines to represent the terms involved. This time we take

$$\bar{A}_{i+1} := s_I - sr_i+1l_{i+1}u_{i+1}u'_{i+1}y_{i+1}y'_{i+1}c_{i+1}x^{n-\lambda_i},$$

$$\bar{B}_{i+1} := dr_i+1l_{i+1}m_{i+1}(su_{i+1}y_{i+1}z_{i+1})^2x^{n-\lambda_i},$$

$$\bar{\Delta}_{i+1} := (\bar{A}_{i+1}, \bar{B}_{i+1}).$$

Then, $\bar{A}_{i+1} = dP_{h_i+1} + (s_1 + td)/2$, which leads to $d_y u_{i+1}l_{i+1}y_{i+1} | \bar{A}_{i+1}$ after a short calculation. We also define

$$\bar{u}_{i+1} := (m_{i+1}, u / u_{i+1}) , \quad \bar{\varepsilon}_{i+1} := sd_y u_{i+1}l_{i+1}y_{i+1}r_{i+1} \bar{u}_{i+1}.$$

By writing $\Delta'_{i+1} = (\bar{G}_{i+1}, u_{i+1}y_{i+1})$, a little calculation gives

$$\bar{\Delta}_{i+1} = sd_y u_{i+1}l_{i+1}y_{i+1}r_{i+1} \bar{u}_{i+1} \Delta'_{i+1}.$$

Note that

$$\frac{\bar{B}_{i+1}}{\bar{\Delta}_{i+1}} = \frac{sdm_{i+1}u_{i+1}y_{i+1}z_{i+1}^2x^{n-\lambda_i}}{d_y \bar{u}_{i+1} \Delta'_{i+1}}.$$

According to Lemma 5.3 there exists two numbers $\bar{y}_{i+2}$ and $\bar{u}_{i+2}$, such that

$$\bar{y}_{i+2} | y/d_y , \quad \bar{u}_{i+2} | u/\bar{w}_{i+1} , \quad \bar{y}_{i+2} \bar{u}_{i+2} \Delta'_{i+1} = \frac{y}{d_y} \frac{u}{\bar{w}_{i+1}} ,$$

and

$$\left( z_{i+1} \bar{y}_{i+2}, \frac{u}{\bar{u}_{i+2} \bar{w}_{i+1}} \right) = 1.$$

From

$$\theta_{h_i+1} = \frac{\bar{A}_{i+1}}{\bar{B}_{i+1}} - \frac{\bar{\alpha}}{r_i+1l_{i+1}m_{i+1}(su_{i+1}y_{i+1}z_{i+1})^2x^{n-\lambda_i}},$$

we apply Lemma 2.2 and find that the next partial quotients are those of the continued fraction expansion of $\bar{A}_{i+1}/\bar{B}_{i+1}$ of length $\bar{p}_{i+1}$, where the parity of $\bar{p}_{i+1}$
is determined by \((-1)^{\hat{h}_{i+1}+1} = \text{sign}(l)\). The next complete quotient is then \(\theta_{j_{i+2}}\), where \(j_{i+2} = h_{i+1} + \hat{p}_{i+1}\) and

\[
\theta_{j_{i+2}} = \frac{r_{i+1}l_{i+1}\beta_{i+1}(su_{i+1}y_{i+1}z_{i+1})^2x^{n-\lambda_i}}{-\text{sign}(l)\alpha\left(\frac{sdm_{i+1}u_{i+1}y_{i+1}z_{i+1}^2}{d_\beta_w_{i+1}A_{i+1}}\right)^2} - \frac{\bar{c}_{i+2}}{B_{i+1}/A_{i+1}},
\]

where \(\bar{c}_{i+2} \equiv -\text{sign}(l)\bar{A}_{i+1}/\bar{A}_{i+1}\) (mod \(B_{i+1}/A_{i+1}\)). The standard choices yield

\[
\theta_{j_{i+2}} = \frac{\omega_n + S_1/d - s\bar{r}_{i+2}\bar{m}_{i+2}\bar{u}_{i+2}\bar{y}_{i+2}\bar{z}_{i+2}\bar{c}_{i+2}x^{2n-\lambda_i}}{\bar{r}_{i+2}\bar{m}_{i+2}(s\bar{u}_{i+2}\bar{y}_{i+2}\bar{z}_{i+2})^2x^{2n-\lambda_i}}.
\]

Since \(s \mid B_{i+1}/\bar{A}_{i+1}\) we get \((\bar{c}_{i+2}, s) = 1\), also \((\bar{z}_{i+2}\bar{y}_{i+2}, u/\bar{u}_{i+2}) = 1\). Hence, the conditions \(C_{i+2}\) are satisfied. Furthermore,

\[
\Psi_{j_{i+2}+1} = \left(\frac{\alpha}{d\bar{A}_{i+1}}\right)\Psi_{h_{i+1}+1}.
\]

The complete quotient in (9-6) satisfies the conditions of Theorem 5.4. Hence, there exists a complete quotient

\[
\theta_{j_{i+3}} = \frac{\omega_n + S_2/d - s\bar{r}_{i+3}\bar{m}_{i+3}\bar{u}_{i+3}\bar{y}_{i+3}\bar{z}_{i+3}\bar{c}_{i+3}x^{\lambda_i+k-n}}{\bar{r}_{i+3}\bar{m}_{i+3}(s\bar{u}_{i+3}\bar{y}_{i+3}\bar{z}_{i+3})^2x^{\lambda_i+k-n}}
\]

with the conditions \(C_{i+3}\) satisfied, and

\[
\Psi_{j_{i+3}+1} = \left(\frac{\beta}{d\bar{A}_{i+2}}\right)\Psi_{j_{i+2}+1} = \left(\frac{\alpha\beta}{d^2\bar{A}_{i+1}\bar{A}_{i+2}}\right)\Psi_{h_{i+1}+1}.
\]

Since \(\epsilon_i = 1\) we have \(\lambda_{i+2} = \lambda_i + k - n\). With appropriate renaming, the complete quotient \(\theta_{j_{i+3}}\) in (9-7) becomes

\[
\theta_{h_{i+2}} = \frac{\omega_n + S_2/d - s\bar{r}_{i+2}\bar{m}_{i+2}\bar{u}_{i+2}\bar{y}_{i+2}\bar{z}_{i+2}\bar{c}_{i+2}x^{\lambda_i+k+2}}{\bar{r}_{i+2}\bar{m}_{i+2}(s\bar{u}_{i+2}\bar{y}_{i+2}\bar{z}_{i+2})^2x^{\lambda_i+k+2}}
\]

with the conditions \(C_{i+2}\) being satisfied. Combining both the \(\epsilon_i = 0\) and \(\epsilon_i = 1\) cases, we have

\[
\Psi_{h_{i+2}+1} = \left(\frac{\alpha^\epsilon_i\beta}{d\bar{A}_{i+1}}\right)\Psi_{h_{i+1}+1} = \left(\frac{\alpha^\epsilon_i+1\beta}{d^2\bar{A}_{i}\bar{A}_{i+1}}\right)\Psi_{h_{i+1}}.
\]

and \(h_{i+2} = h_{i} + p_{i} + p_{i+1}\).

Our next step will be to investigate the number of partial quotients in the period of the continued fraction expansion of \(\omega_n\).
10. Determining the number of partial quotients

By defining, \( f_0 := h_0 \) and \( f_{i+1} := p_i \) for \( i \geq 1 \) we find

\[
(10-1) \quad h_i = \sum_{k=0}^{i} f_k.
\]

The value of \( f_i \) is dependent only upon the set

\[
Z_i := \{l_i, m_i, r_i, s_i, s_i', z_i, z_i', u_i, u_i', y_i, y_i', c_i, L_i, x\},
\]

where \( L_i := \lambda_i \pmod{\mu} \). Moreover, if \( Z_i = Z_j \) then \( f_i = f_j \). There are only finitely many distinct sets \( Z_i \); we denote the total number of distinct sets by \( Z \). For a fixed \( t \), there are precisely \( tk \) values of \( i \in \{1, \ldots, 2tn - 1\} \), where \( \varepsilon_i = 1 \). Let \( i_1, \ldots, i_k \) represent these points. Then

\[
i_h = \begin{cases} \frac{2nh}{k} - 3 & \text{if } k \nmid 2nh, \\ \left\lfloor \frac{2nh}{k} \right\rfloor - 2 & \text{if } k \mid 2nh. \end{cases}
\]

From (10-1),

\[
h_{2tn-1} = \sum_{i=0}^{i-1} f_i + \sum_{h=1}^{i-1} f_h + \sum_{h=1}^{i-1} \sum_{i=i_h+1}^{i_h+1} f_i.
\]

The number of summands in \( \sum_{i_h+1}^{i_h+1} f_i \) is \( i_{h+1} - 1 - (i_h + 1) + 1 \geq \left\lfloor \frac{2n}{\kappa} \right\rfloor - 2 \). Hence the distance between \( i_h \) and \( i_{h+1} \) can be made arbitrarily large. But \( Z \) is independent of \( n \), which by the box principle means that for large enough \( n \), there exists \( \rho_h \) and \( \tau_h \) such that

\[
Z_{i_h+\tau_h} = Z_{i_h+\tau_h+\rho_h} \quad i_h + \tau_h + \rho_h \leq i_{h+1} - 1 \quad 1 \leq \tau_h, \rho_h \leq Z.
\]

Now we examine \( Z_{i_h+\tau_h+j\rho_h} \), where \( \kappa + i_h + \tau_h + j\rho_h \leq i_{h+1} - 1 \) and \( 0 \leq \kappa < \rho_h \). Since \( \varepsilon_{i_h+\tau_h} = 0 \) and \( \varepsilon_{i_h+\tau_h+\rho_h} = 0 \) we have

\[
Z_{i_h+\tau_h+1} = Z_{i_h+\tau_h+\rho_h+1} \quad \text{provided } i_h + \tau_h + \rho_h + 1 \leq i_{h+1} - 1.
\]

By induction,

\[
Z_{i_h+\tau_h+\kappa} = Z_{i_h+\tau_h+\kappa+j\rho_h} \quad \text{provided } i_h + \tau_h + \kappa + j\rho_h \leq i_{h+1} - 1,
\]

which implies \( f_{i_h+\tau_h+\kappa} = f_{i_h+\tau_h+\kappa+j\rho_h} \). Define

\[
u := \left\lfloor \frac{\left\lfloor \frac{2n}{\kappa} \right\rfloor - \tau_h - \rho_h}{\rho_h} \right\rfloor - 1.
\]
It is straightforward that \( i_h + \tau_h + \kappa + \nu \rho_h < i_{h+1} \). Then

\[
\sum_{i=i_h+1}^{i_{h+1}-1} f_i = \sum_{j=1}^{\tau_h-1} f_{i_h+j} + \sum_{\kappa=0}^{\rho_h-1} \left( \sum_{j=0}^{\nu} f_{i_h+\tau_h+\kappa+j\rho_h} \right) + \sum_{j=i_h+\tau_h+\rho_h+\nu \rho_h}^{i_{h+1}-1} f_j.
\]

The number of terms in \( \sum_{j=i_h+\tau_h+\rho_h+\nu \rho_h}^{i_{h+1}-1} f_j \) is \( \leq \rho_h + 2 \). Hence,

\[
(10-2) \sum_{i=i_h+1}^{i_{h+1}-1} f_i = \left[ \frac{2n - k \tau_h - k \rho_h}{k \rho_h} \right] \zeta_h + \xi_h,
\]

where \( \zeta_h, \xi_h \) are independent of \( n \).

We now take \( \rho = \prod_{i=1}^{k} \rho_h \) and \( w = \text{lcm} \{ k, \mu, \rho \} \) both of which are independent of \( n \). We write \( n = w \gamma + \phi \), where \( 0 \leq \phi < w \). From the original set \( I_n \), we now wish to consider the following subset, \( I_{\nu, \phi} = \{ n \in I_n : n \equiv \phi \pmod{w} \} \). Without loss of generality, we may suppose that \( n \in I_{\nu, \phi} \). Consequently,

\[
\left[ \frac{2n - k \tau_h - k \rho_h}{k \rho_h} \right] = 2 \gamma \frac{w}{k \rho_h} + \left[ \frac{2 \phi - k \tau_h - k \rho_h}{k \rho_h} \right].
\]

Thus, the sum (10-2) becomes,

\[
\sum_{i=i_h+1}^{i_{h+1}-1} f_i = 2 \zeta_h \gamma \frac{w}{k \rho_h} + \zeta_h \left[ \frac{2 \phi - k \tau_h - k \rho_h}{k \rho_h} \right] + \xi_h,
\]

which means we can now write \( h_{2n-1} \) as

\[
h_{2n-1} = \sum_{i=0}^{i_{h+1}-1} f_i + \sum_{h=1}^{t \kappa - 1} f_{i_h} + 2 \gamma \sum_{h=1}^{t \kappa - 1} \zeta_h \frac{w}{k \rho_h} + \sum_{h=1}^{t \kappa - 1} \left( \zeta_h \left[ \frac{2 \phi - k \tau_h - k \rho_h}{k \rho_h} \right] + \xi_h \right).
\]

Now, let

\[
x_t = 2 \sum_{h=1}^{t \kappa - 1} \zeta_h \frac{w}{k \rho_h}
\]

\[
y_t = \sum_{h=1}^{t \kappa - 1} \left( \zeta_h \left[ \frac{2 \phi - k \tau_h - k \rho_h}{k \rho_h} \right] + \xi_h \right) + \sum_{i=0}^{i_{h+1}-1} f_i + \sum_{h=1}^{t \kappa - 1} f_{i_h},
\]

both of which are integers and independent of \( \gamma \). Then

\[
h_{2n-1} = \gamma x_t + y_t = \left( (n - \phi)/w \right) x_t + y_t = a_t n + b_t,
\]

where \( a_t, b_t \) are both rational numbers independent of \( n \) for all \( n \in I_{\nu, \phi} \).
This shows that the length of the expansion up to $h_{2nt-1}$ is linear in $n$. It remains to show that there exists some $h_{2nt-1}$, where $Q_{h_{2nt-1}} = 1$, and if $Q_j = 1$ then $j = h_{2nt-1}$ for some $t$ independent of $n$.

11. Finding an element of norm 1

We now examine the product of the elements in the expansion. By Theorem 9.1 and (9-2),

$$\Psi_{h_{2n-1}+1} = \frac{(\alpha \beta)^{\sum_{j=1}^{t-1} t_j} \alpha^{\sum_{j=1}^{t-1} t_j}}{d^{2n-1} \prod_{j=0}^{2n-2} \Delta_j}.$$ 

So,

$$\Psi_{h_{2n-1}+1} = \frac{(\alpha \beta)^{nt} \alpha^{kt}}{d^{2n-1} \prod_{j=0}^{2n-2} \Delta_j} \quad \text{and} \quad |N(\Psi_{h_{2n-1}+1})| = Q_{h_{2n-1}}.$$

Hence,

$$Q_{h_{2n-1}} \left( d^{2n-1} \prod_{j=0}^{2n-2} \Delta_j \right)^2 = |N((\alpha \beta)^{nt}) N(\alpha^{kt})| = |N((\alpha \beta)^{nt} N(\alpha^k)|^t.$$

as well as

$$Q_{h_{2n-1}} \left( d^{2n-1} \prod_{j=0}^{2n-2} \Delta_j \right)^2 = |N((\alpha \beta)^{nt} N(\alpha^k)|.$$

Thus, $\sqrt{Q_{h_{2n-1}} / (Q_{h_{2n-1}})^t} \in Q$. Since $\lambda_{2j-1} \equiv jk \pmod{n}$ we have that $\lambda_{2n-1} = 0$ for positive $i$. Hence, $\lambda_{2n-1} = \lambda_{2n-1} = 0$ and so

$$Q_{h_{2n-1}} = l_{2n-1}m_{2n-1}r_{2n-1} (su_{2n-1}z_{2n-1}y_{2n-1})^2,$$

$$Q_{h_{2n-1}} = l_{2n-1}m_{2n-1}r_{2n-1} (su_{2n-1}z_{2n-1}y_{2n-1})^2.$$

Thus,

$$\sqrt{\frac{l_{2n-1}m_{2n-1}r_{2n-1}}{(l_{2n-1}m_{2n-1}r_{2n-1})^2}} \in Q.$$

Since $l_{2n-1}, m_{2n-1}, r_{2n-1}$ are each squarefree and relatively prime, if $2 \mid t$ then $\sqrt{l_{2n-1}m_{2n-1}r_{2n-1}} \in Q$, which implies $l_{2n-1}m_{2n-1}r_{2n-1} = 1$. Conversely, if $2 \nmid t$, then $l_{2n-1}m_{2n-1}r_{2n-1} = l_{2n-1}m_{2n-1}r_{2n-1}$.

Now, we construct an element of norm 1. Put

$$\varepsilon = \begin{cases} 
1 & \text{if } l_{2n-1}m_{2n-1}r_{2n-1} \neq 1, \\
0 & \text{if } l_{2n-1}m_{2n-1}r_{2n-1} = 1,
\end{cases}$$
and

\[ \Gamma = \frac{\Psi_{n_{2n-1}+1}^{1+\varepsilon}}{(l_{2n-1}m_{2n-1}r_{2n-1})^\varepsilon (s_{2n-1}t_{2n-1}r_{2n-1})^{1+\varepsilon}}. \]

Then it is easily shown that \( N(\Gamma) = 1. \)

**Lemma 11.1.** If \( D_n = F_n^2 D'_n, \) where \( D'_n \) is squarefree, then \( su_1 y_1 z_1 | F_n. \)

The lemma and above results imply that we have \( V_1, Y_1 \in \mathbb{Z} \) such that

\[ \Gamma = V_1 + Y_1 \frac{F_n}{s_{2n-1}t_{2n-1}r_{2n-1}} \theta_n, \]

where \( \mathbb{Z}[\theta_n] \) is the maximal order of \( \mathbb{Q}(\sqrt{D'_n}) \) and \( D'_n \) is the squarefree kernel of \( D_n, \) that is \( D_n = F_n^2 D'_n. \) If we now define \( V_j \) and \( Y_j \) by

\[ \left( V_j + Y_j \frac{F_n}{s_{2n-1}t_{2n-1}r_{2n-1}} \theta_n \right) = \Gamma^j. \]

Then \( Y_j / Y_1 \) is the Lucas function, \( (\Gamma^j - \Gamma^j)/(\Gamma - \Gamma). \) Since \( \Gamma \Gamma = 1, \) there must exist some minimal positive \( p \) such that \( su_{2n-1}t_{2n-1}r_{2n-1} | Y_p. \) Putting \( t = g := (1 + \varepsilon) p \) we get

\[ \Psi_{n_{2n-1}+1} = \Gamma^p \sqrt{l_{2n-1}m_{2n-1}r_{2n-1}t_{2n-1}y_{2n-1}z_{2n-1}^2}. \]

If \( \varepsilon = 1 \) then \( 2 | g \) implies that \( l_{2n-1}m_{2n-1}r_{2n-1} = 1. \) If \( \varepsilon = 0 \) then

\[ l_{2n-1}m_{2n-1}r_{2n-1} = 1, \]

so \( l_{2n-1}m_{2n-1}r_{2n-1} = 1. \) Since \( \Gamma^p \in \mathbb{Z}[\theta_n], \) this implies \( su_{2n-1}t_{2n-1}r_{2n-1} = 1. \) Hence \( |N(\Psi_{n_{2n-1}+1})| = 1, \) which means that \( Q_{n_{2n-1}+1} = 1. \) The values of \( p \) depend only on \( su_{2n-1}t_{2n-1}r_{2n-1}, \) which divides \( suyz. \) Thus, there can only be a finite number of possible values for \( p. \)

Conversely, one can also show (see [Patterson 2003, Chapter 25]) that such a solution is either fundamental or the square of the fundamental solution, although this is superfluous in showing that \( D_n \) is a kreeper.

**12. Returning to the regular continued fraction expansion**

Up to now we have determined

\[ \omega_n = \{ a_0, \ldots, a_{h_0-1}, b_1, a_{h_0+1}, \ldots, a_{b_1-1}, b_2, \ldots, a_{h_{2n-1}-1}, 2a_0 - t \}, \]

but in this evaluation we never insisted that \( b_i \geq 1. \) In other words this expansion might not correspond to the regular continued fraction expansion of \( \omega_n. \) This is equivalent to saying that the expansions of \( A_i / B_i \) might have an initial nonpositive partial quotient.
Proposition 12.1. In the expansion of \( \omega_n \) given by the earlier procedure, the number of nonpositive partial quotients is bounded independently of \( n \).

Proof. See [Patterson 2003].

This proposition means that there are only finitely many partial quotients that need to be altered in order to find the regular continued fraction expansion of \( \omega_n \).

The removal of nonpositive partial quotients is covered in [Dirichlet 1999]. There it is shown that any negative partial quotient can be moved to the left in the continued fraction expansion. In our situation, we discover that either the negative partial quotient disappears easily or we are left with

\[
[ b_{i-1}, -1, 1, u, v, \ldots ] = [ b_{i-1} - 2 - u, 1, v - 1, \ldots ],
\]

where \( u > 0 \) and is bounded independently of \( n \), and \( b_{i-1} = [A_{i-1}/B_{i-1}] \). When

\[
\theta_{h_{i-1}} = \frac{\omega_n + S_2/d - E_{i-1}x^{\lambda_i}/d}{E_{i-1}x^{\lambda_i}} = \frac{A_{i-1}}{B_{i-1}} + e
\]

we have \( \lambda_i \ll 1 \), implies \( [A_{i-1}/B_{i-1}] > x^{n-J_1} \), where \( J_1 \) is bounded independently of \( n \). Consequently, \( b_{i-1} - 2 - u > 0 \) for all sufficiently large \( n \). The other case follows similarly.

Finally, we note that \( a_{h_{i+j}} \) with \( j > 0 \) can not be the end of the period since \( Q_{h_{i+j}} > 1 \). For sufficiently large \( n \), each \( b_j \) which is not some \( b_{h_{2tn-1}} \) satisfies \( b_j < x^n \) since \( B_i \geq x \).

In conclusion, we have shown that

\[
lp(\omega_n) = h_{2n^{k-1}} + c_n = a_g n + b_g + c_n = a_g n + b'_g,
\]

where \( c_n \in \mathbb{Z} \) can be bounded independently of \( n \). Then \( a_g, b'_g \) are rational numbers bounded independently of \( n \).

Hence there must exist an infinitude of values of \( n \in I \) such that

\[
lp(\omega_n) = an + b,
\]

where \( a, b \in \mathbb{Q} \) and are fixed independently of \( n \). This completes the proof of Theorem 1.2.

References


COMBINATORIAL RIGIDITY IN CURVE COMPLEXES AND MAPPING CLASS GROUPS

KENNETH J. SHACKLETON

In all possible cases, we prove that local embeddings between two curve complexes whose complexities do not increase from domain to codomain are induced by surface homeomorphism. This is our first main result. From this we can deduce our second, a strong local co-Hopfian property for mapping class groups.

Introduction

The curve complex $\mathcal{C}(\Sigma)$ associated to a surface $\Sigma$ was introduced by Harvey [1981] to encode the large scale geometry of Teichmüller space, and ultimately help decide the nonarithmeticity of the surface mapping class groups. It was later to play a central role in the proof by Minsky, Brock and Canary of Thurston’s ending lamination conjecture [Brock et al. 2004].

We start by defining the curve complex, and throughout our surfaces will be compact, connected and orientable. We say that a simple loop on $\Sigma$ is trivial if it bounds a disc and peripheral if it bounds an annulus whose other boundary component belongs to $\partial \Sigma$. A curve on $\Sigma$ is a free homotopy class of a nontrivial and nonperipheral simple loop and we denote the set of these by $X(\Sigma)$. The intersection number of two curves $\alpha, \beta \in X(\Sigma)$, denoted $\iota(\alpha, \beta)$, is defined equal to $\min \{|a \cap b| : a \in \alpha, b \in \beta\}$. We say that two curves intersect minimally if they intersect once or they intersect twice with zero algebraic intersection and refer to either as the type of minimal intersection. We will later define the complexity of $\Sigma$, denoted $\kappa(\Sigma)$, as equal to the maximal number of distinct and disjoint curves that can be realised simultaneously.

When $\kappa(\Sigma) \geq 2$, the curve graph is the graph whose vertex set is $X(\Sigma)$ and we deem two distinct curves to span an edge if and only if they can be realised disjointly in $\Sigma$. When $\kappa(\Sigma) = 1$, we say that two distinct curves are joined by an edge if and only if they intersect minimally. The curve complex associated to $\Sigma$ is the curve graph when $\kappa(\Sigma) = 1$, making it isomorphic to a Farey graph, and the

\begin{itemize}
  \item \textbf{MSC2000}: primary 57M50; secondary 20F38.
  \item \textbf{Keywords}: curve complex, mapping class group.
  \item Supported by an EPSRC PhD scholarship.
\end{itemize}
flag simplicial complex whose 1-skeleton is the curve graph when $\kappa(\Sigma) \geq 2$. In the latter case, $\mathcal{C}(\Sigma)$ has simplicial dimension precisely $\kappa(\Sigma) - 1$.

For each curve $\alpha$ we denote by $X(\alpha)$ the set of all curves on $\Sigma$ distinct and disjoint from $\alpha$, that is the vertex set of the link of $\alpha$. This link is always connected whenever $\kappa(\Sigma)$ is at least three, and whenever $\kappa(\Sigma)$ is two any two elements of $X(\alpha)$ may be “chain-connected” by a finite sequence of curves in which any two consecutive curves have minimal intersection.

In this paper, we shall be discussing embeddings between two curve complexes whose complexities do not increase from domain to codomain and we shall find that these are all induced by surface homeomorphism, so long as we place a necessary but consistent hypothesis in one sporadic case. The argument we give is by an induction on complexity and requires little more than the connectivity of links in the curve complex over and above this. As such, our approach does not discriminate in terms of the topological type of a surface. Moreover, we actually only require the local injectivity of an embedding and we shall say more on this towards the end of this section.

All told, this generalises the automorphism theorem of Ivanov for surfaces of genus at least two, a proof of which is sketched in [Ivanov 1997, §2] and extended in [Korkmaz 1999] to all but the two-holed torus, and of Luo [2000], settled or proven in all cases. Making use of their combined result, Margalit [2004] establishes the analogue for automorphisms of another important surface complex called the pants complex. There are analogues for other surface complexes; see [Schmutz Schaller 2000] as one example.

**Theorem 1.** Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces of positive complexity such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$, and that when they have equal complexity at most three they are homeomorphic or one is the three-holed torus. Then, any simplicial embedding from $\mathcal{C}(\Sigma_1)$ to $\mathcal{C}(\Sigma_2)$ (preserving the separating type of each curve when the two surfaces are homeomorphic to the two-holed torus) is induced by a surface homeomorphism.

This covers all possibilities. We remind ourselves that there exist isomorphisms between the curve complex of the closed surface of genus two and the six-holed sphere, the two-holed torus and the five-holed sphere and finally the one-holed torus and the four-holed sphere and that there exists an automorphism of the curve complex associated to the two-holed torus that sends a nonseparating curve to an outer curve (see [Luo 2000] for more details). These are examples of embeddings not induced by a surface homeomorphism. Finally, we point out that there exist embeddings on curve complexes with complexity increasing from domain to codomain not induced by a surface embedding: Easy examples are provided by taking some proper subsurface $\Sigma_1$ of $\Sigma_2$, and modifying the induced embedding.
on curve complexes by instead taking just one curve on $\Sigma_1$ to a curve on $\Sigma_2$ outside of $\Sigma_1$.

Among other things, Theorem 1 completes one study of a particular class of self-embedding, initiated by Irmak. This class comprises the superinjective maps, and by definition each preserves the nonzero intersection property of a pair of curves. Irmak [2004b] first showed that a superinjective self-map is induced by a surface homeomorphism provided the surface is closed and of genus at least three, then [2006] extended this result to nonclosed surfaces of genus at least three and surfaces of genus two with at least two holes, and to the remaining two types of genus two surface [2004a]. Following a now standard strategy, first set out by Ivanov, this holds consequences for the mapping class groups of the corresponding surfaces.

The mapping class group $\text{Map} \, \Sigma_1$ is the group of all self-homeomorphisms of the surface $\Sigma$, up to homotopy. This is sometimes known as the extended mapping class group, for it contains the group of orientation preserving mapping classes as an index two subgroup. Some of its other subgroups, in particular the Johnson kernel and the Torelli group, are of wide interest; see, respectively, [Brendle and Margalit 2004] and [Farb and Ivanov 2005], and references contained therein.

The mapping class group has a natural simplicial action on the curve complex, determined by first lifting a curve to a representative loop and then taking the free homotopy class of the image under a representative homeomorphism. The kernel of this action, $\ker \Sigma$, is almost always trivial: the only exceptions lie in low complexity, where this kernel is isomorphic to $\mathbb{Z}_2$ and generated by the hyperelliptic involution when $\Sigma$ is the one-holed torus, the two-holed torus, or the closed surface of genus two or isomorphic to $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ and generated by two hyperelliptic involutions when the four-holed sphere; this was proved in [Birman 1974; Viro 1972]. For a detailed account of the mapping class group and its subgroups, see [Ivanov 1992] as one place to start.

Theorem 1 implies the following strong co-Hopfian property for mapping class groups. Among other things Theorem 2 has some familiar consequences, namely it follows that the commensurator group of a mapping class group is isomorphic to the same mapping class group and that the outer automorphism group of a mapping class group is trivial. It also follows that mapping class groups do not admit a faithful action on another curve complex of no greater dimension.

**Theorem 2.** Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$ and at least two, and that whenever they both have complexity equal to three they are homeomorphic, though not to the closed surface of genus two, or one is the three-holed torus and that when they both have complexity two they are homeomorphic to the five-holed sphere. Suppose that $H$ is a finite index subgroup of the mapping class group $\text{Map} \, \Sigma_1$. Then, every injection of $H$ into $\text{Map} \, \Sigma_2$ is the restriction of an inner automorphism of $\text{Map} \, \Sigma_1$. 
The existence of such a homomorphism is understood to imply the two surfaces are equal. Theorem 2 is a generalisation of [Ivanov and McCarthy 1999, Theorem 4], which deals with injections defined on mapping class groups associated to surfaces of positive genus.

The combined superinjectivity theorem implies Theorem 2 when the two surfaces under consideration are homeomorphic and have genus at least three, or genus at least two and one hole, and Irmak [2004a] describes a noninner automorphism for the closed surface of genus two. Bell and Margalit [2005] extend this to spheres with at least five holes, and Behrstock and Margalit [2006] to genus one surfaces with at least three holes in addition to finding a commensurator for the mapping class group of the two-holed torus not induced by an inner automorphism. The remaining cases, namely the mapping class group of the four-holed sphere and of the one-holed torus, also have noninner injections on finite index subgroups, as both are virtually free groups. We remark the braid groups on at least four strands, modulo centre have the co-Hopfian property; see [Bell and Margalit 2006].

The general approach we need for Theorem 2 follows that given by Ivanov, translating an injection on a finite index subgroup to an embedding on curve complexes. This is now a well-established strategy on which we have nothing to add, and a thorough account can be found in both [Bell and Margalit 2005] and [Irmak 2004b].

Though all our arguments are phrased in terms of embeddings, they only ever need the simplicial and local injectivity properties of such maps. We can therefore record the following generalisation of Theorem 1, the first of two main results. Recall that a star is the union of all edges incident on a common vertex.

**Theorem 3.** Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces of positive complexity such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$, and that when they have equal complexity at most three they are homeomorphic or one is the three-holed torus. Then, any simplicial map from $\mathcal{C}(\Sigma_1)$ to $\mathcal{C}(\Sigma_2)$ injective on every star (and preserving the separating type of each curve when both surfaces are homeomorphic to the two-holed torus) is induced by a surface homeomorphism.

Again, this covers all possibilities. We remark that proving a local embedding is induced by a surface homeomorphism would appear the most direct way of seeing that it must also be a global embedding. Furthermore, we conjecture that the pants complex also exhibits such local-to-global rigidity.

From Theorem 3 we can deduce, using a careful application of Ivanov’s strategy, the following local version of Theorem 2. This is one interpretation of local injectivity for mapping class groups, and a proof is completed in [Shackleton 2005, §3.3]. Among other things, it follows that a self-homomorphism of a mapping class group injective on every curve stabiliser is the restriction of an inner automorphism.
Theorem 4. Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$ and at least two, and that whenever they both have complexity equal to three they are homeomorphic, though not to the closed surface of genus two, or one is the three-holed torus and that whenever they both have complexity two they are homeomorphic to the five-holed sphere. Suppose that $H$ is a finite index subgroup of the mapping class group $\text{Map} \, \Sigma_1$. Then, every homomorphism of $H$ into $\text{Map} \, \Sigma_2$ injective on every curve stabiliser in $H$ is the restriction of an inner automorphism of $\text{Map} \, \Sigma_1$.

Note once more that the existence of such a homomorphism is understood to imply the two surfaces are equal.

Investigations into arbitrary homomorphisms from a mapping class group associated to a closed surface of genus at least one to another mapping class group associated to a closed surface of smaller genus have been made Harvey and Korkmaz [2005], who found that every such homomorphism has finite image. Their approach seems to make essential use of the existence of torsion in mapping class groups and, as mapping class groups are virtually torsion free, it would be of some interest to find a way around this so as to consider finite index subgroups.

1. Embeddings between curve complexes

For any compact, connected and orientable surface $\Sigma$ the complexity $\kappa(\Sigma)$ of $\Sigma$ is defined to be equal to $3\text{genus} \, \Sigma + |\partial \Sigma| - 3$. This is perhaps nonstandard, since complexity is often taken to be equal to the simplicial dimension of the curve complex, but the additivity of $\kappa$ best suits our induction argument. By way of example, the one-holed torus and the four-holed sphere are the only surfaces of complexity one, the two-holed torus and the five-holed sphere are the only surfaces of complexity two, and the closed surface of genus two, the three-holed torus and the six-holed sphere are the only surfaces of complexity three. On occasion we refer to these as the low complexity surfaces.

We shall abuse notation slightly by viewing each curve as a vertex, as a class of loops, or as a simple loop already realised on $\Sigma$. Our interpretation will be apparent from the context. We say that a curve is separating if its complement is not connected, and otherwise say it is nonseparating. We say that a curve is an outer curve if it is separating and if it bounds a two-holed disc (equivalently, a three-holed sphere). These are usually known as boundary curves in the literature, but here we need to avoid confusing them with the components of $\partial \Sigma$. A multicurve on $\Sigma$ is a collection of distinct and disjoint curves, and a pants decomposition of $\Sigma$ is a maximal multicurve. A pair of pants in $\Sigma$ is an essential subsurface homeomorphic to a compact three-holed sphere.
We say that two curves in a pants decomposition $P$ are adjacent in $P$ if they appear in the boundary of a three-holed sphere complementary to $P$. This is slightly unfortunate terminology that seems to be a long way to becoming standard; we sincerely hope that any confusion between adjacency in the curve complex and adjacency in a pants decomposition will be obviated by the context.

The structure of our argument is broadly as follows. We establish a short list of topological properties verified by any embedding on curve complexes from which we easily deduce, among other things, that the existence of such an embedding implies the two surfaces have equal complexity and then, with more work, almost always means the two surfaces under consideration are homeomorphic. For the time being, we refer to embeddings between two apparently distinct curve complexes as cross-embeddings. Dealing with embeddings in low complexity typically requires individual arguments and it therefore streamlines our work if we do this separately, as we do in Lemma 13. The proof of Theorem 1 is then completed by an induction on complexity, where we cut the surface along a curve. As embeddings behave well on the topological type of a curve, the resulting surfaces are again homeomorphic. For the induction argument to pass through complexity one (sub)surfaces, we will need to know that embeddings preserve minimal intersection.

We start by showing, in turn, that embeddings send pants decompositions to pants decompositions, they preserve a form of small intersection and they preserve adjacency and nonadjacency in a pants decomposition.

**Lemma 5.** Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$. Then, any embedding $\phi$ from $\mathcal{C}(\Sigma_1)$ to $\mathcal{C}(\Sigma_2)$ sends pants decompositions to pants decompositions.

**Proof.** This follows for complexity reasons and because $\phi$ is simplicial and injective. \qed

To make sense of the following lemma, we must define what we mean by the subsurface of $\Sigma$ filled by two curves $\alpha$ and $\beta$. Letting $N(\alpha \cup \beta)$ denote a closed regular neighbourhood of $\alpha \cup \beta$ in $\Sigma$, we augment $N(\alpha \cup \beta)$ by taking its union with all the complementary discs whose boundary is contained in $N(\alpha \cup \beta)$ and all the complementary annuli with one boundary component in $\partial \Sigma$ and the other in $N(\alpha \cup \beta)$. The resulting subsurface of $\Sigma$ is well defined up to isotopy and is what we mean by the subsurface filled by $\alpha$ and $\beta$. Whenever a third curve enters the subsurface filled by two curves, it must intersect at least one of these two curves.

**Lemma 6.** Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$. Let $\phi$ be any embedding from $\mathcal{C}(\Sigma_1)$ to $\mathcal{C}(\Sigma_2)$ and let $\alpha, \beta$ be any two curves in $\Sigma_1$ that fill either a four-holed sphere or a one-holed torus. Then, $\phi(\alpha)$ and $\phi(\beta)$ fill either a four-holed sphere or a one-holed torus in $\Sigma_2$. 


Proof. Let $Q$ be any maximal multicurve in $\Sigma_1$ such that each curve is disjoint from both $\alpha$ and $\beta$. For complexity reasons, $\phi(Q)$ is a maximal multicurve disjoint from both $\phi(\alpha)$ and $\phi(\beta)$. In particular, as $\phi$ is injective and simplicial so $\phi(\alpha)$ and $\phi(\beta)$ must together fill either a four-holed sphere or a one-holed torus. □

We say that two curves have small intersection if they together fill either a four-holed sphere or a one-holed torus, and refer to either as the type of the small intersection. Any two curves that intersect minimally also have small intersection, but the converse does not hold.

**Lemma 7.** Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$. Let $P$ be any pants decomposition of $\Sigma_1$ and let $\phi$ be any embedding from $\mathcal{C}(\Sigma_1)$ to $\mathcal{C}(\Sigma_2)$. Then, any two curves adjacent in $P$ are sent by $\phi$ to two curves adjacent in $\phi(P)$ and any two curves in $P$ that are not adjacent in $P$ are sent by $\phi$ to two curves not adjacent in $\phi(P)$.

Proof. The first part follows from Lemma 6: For any two curves $\alpha_1$ and $\alpha_2$ adjacent in $P$, there exists a curve $\delta$ having small intersection with both and disjoint from every other curve in $P$. This is preserved under $\phi$ and so $\phi(\alpha_1)$ and $\phi(\alpha_2)$ are adjacent in $\phi(P)$.

Similarly, if two curves $\alpha_1, \alpha_2$ are not adjacent in $P$ we can find two disjoint curves $\delta_1, \delta_2$ such that $\delta_1$ has small intersection with $\alpha_1$ but is disjoint from $\alpha_2$ and $\delta_2$ has small intersection with $\alpha_2$ but is disjoint from $\alpha_1$ and both $\delta_1, \delta_2$ are disjoint from every other curve in $P$. If $\phi(\alpha_1)$ and $\phi(\alpha_2)$ are adjacent in $\phi(P)$ then $\phi(\delta_1)$ and $\phi(\delta_2)$ must intersect. As $\phi$ is simplicial, this is a contradiction. □

The import of Lemma 5, Lemma 6 and Lemma 7 is perhaps best understood by associating to a pants decomposition $P$ a certain graph. The vertices of this graph are the curves in $P$, and any two distinct vertices span an edge if and only if they correspond to adjacent curves in $P$. Lemma 7 not only tells us that any embedding $\phi$ induces a map between adjacency graphs, but that this map is actually an isomorphism. Cut points in the graph correspond to nonouter separating curves, and noncut points correspond to outer or nonseparating curves.

This graph, and the ideas bound by Lemma 7, were independently and simultaneously discovered by Behrstock and Margalit. Their approach, published in a joint paper [2006], and the arguments they give will deal with all superinjective maps for two homeomorphic surfaces of complexity at least three. From this they also deduce that the commensurator group of a mapping class group is isomorphic to the same mapping class group. We both refer to such a graph as an adjacency graph.

We can just as well speak of an adjacency graph associated to a multicurve $Q$, in which the vertices again correspond to the curves in $Q$ and any two vertices are declared adjacent if their corresponding curves border a common pair of pants in
Figure 1. A codimension 1 multicurve, with its adjacency graph.

the surface complement of $Q$. There is a subtle point to be made here, namely that the complementary graph of a vertex in a pants adjacency graph will not in general be the adjacency graph of the multicurve that results by removing the corresponding curve from the pants decomposition. It will however be the adjacency graph that results from cutting the surface along this curve. By way of example, on removing a curve $\alpha$ from a pants decomposition $P$ any curves that together bound a complementary four-holed sphere will not necessarily be adjacent in the adjacency graph of $P - \{\alpha\}$. (See Figure 1 for an example.) This observation will be important later when we come to look at outer curves. It does however hold that a curve complex embedding induces an isomorphism between multicurve adjacency graphs.

**Lemma 8.** Suppose that $\Sigma_1$ and $\Sigma_2$ are compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$. Let $Q$ be any multicurve of $\Sigma_1$ and let $\phi$ be any embedding from $\mathcal{C}(\Sigma_1)$ to $\mathcal{C}(\Sigma_2)$. Then, $\phi$ induces an isomorphism from the adjacency graph of $Q$ to the adjacency graph of $\phi(Q)$.

**Proof.** We make use of Lemma 7. Extend $Q$ to a pants decomposition $P$ of $\Sigma_1$. If two curves are adjacent in $Q$ then they either border a pair of pants with a third curve from $Q$ or they border a pair of pants meeting $\partial \Sigma$. This remains so in $P$, and is preserved on applying $\phi$. To show nonadjacency is preserved, consider any two curves not adjacent in $Q$ and arrange for them to be nonadjacent in $P$. This is preserved under $\phi$. 

□
As embeddings between curve complexes induce isomorphisms on adjacency graphs and graph isomorphisms send cut points to cut points, so embeddings must send nonouter separating curves to nonouter separating curves.

**Lemma 9.** Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$. Then, any embedding $\phi$ from $\mathcal{C}(\Sigma_1)$ to $\mathcal{C}(\Sigma_2)$ sends nonouter separating curves to nonouter separating curves.

We use the adjacency graph to distinguish between nonseparating and outer curves.

**Lemma 10.** Suppose $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$ and that whenever they have equal complexity at most three they are homeomorphic and not the two-holed torus. Let $\phi : \mathcal{C}(\Sigma_1) \to \mathcal{C}(\Sigma_2)$ be any embedding. Then, $\phi$ takes nonseparating curves to nonseparating curves.

**Proof.** We note that the $\phi$-image of a nonseparating curve can never be a nonouter separating curve, for otherwise we see a noncut point sent to a cut point in some pants adjacency graph. Suppose that $\alpha$ is a nonseparating curve in $\Sigma_1$. When $\kappa(\Sigma_1)$ is at least four we can find a pants decomposition $P$ extending $\alpha$ in which $\alpha$ corresponds to a vertex in the adjacency graph of $P$ of valence three or four. As $\phi$ induces an isomorphism on the adjacency graph, so $\phi(\alpha)$ must have the same valence. As outer curves only ever correspond to vertices of valence at most two, so $\phi(\alpha)$ can only be nonseparating.

With the exception of the two-holed torus, all cases in which $\Sigma_1$ has complexity at most two hold since there is only ever one type of curve. In complexity three, when $\Sigma_1$ is the six-holed sphere our claim holds vacuously and when $\Sigma_1$ is the closed surface of genus two our claim follows from Lemma 9 by noting that every pants decomposition contains at most one separating curve.

The only nontrivial case in low complexity is that of $\Sigma_1$ and $\Sigma_2$ both homeomorphic to the three-holed torus. In which case, there are only two pants adjacency graphs, up to isomorphism, but three different pants decompositions, up to the action of the mapping class group. For this reason, we need to argue differently: If there is a nonseparating curve sent by $\phi$ to an outer curve, then there is an outer curve $\alpha$ sent by $\phi$ to a nonseparating curve. To see this, extend this nonseparating curve to a pants decomposition containing a nonouter separating curve. By appealing to Lemma 9, we see that the third curve in this pants decomposition will suffice. Now extend $\alpha$ to a second pants decomposition containing two nonseparating curves $\delta_1$ and $\delta_2$. The $\phi$-image of at least one of these, say $\delta_1$, is again a nonseparating curve. Choose any two disjoint curves $\gamma_1, \gamma_2$ in $\Sigma_1$ that have small intersection with $\delta_1$ and $\alpha$ but disjoint from $\alpha$ and $\delta_1$, respectively. Now $\phi(\delta_1)$ and $\phi(\alpha)$ border a common pair of pants in $\Sigma_2$ invaded by $\phi(\gamma_1)$ and $\phi(\gamma_2)$. We see that
Lemma 11. Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$, and that whenever they have equal complexity at most three they are homeomorphic and not the two-holed torus. Then, any embedding $\phi$ from $\mathcal{E}(\Sigma_1)$ to $\mathcal{E}(\Sigma_2)$ sends outer curves to outer curves.

Proof. We note that this holds vacuously when $|\partial \Sigma_1|$ is at most one. In any case, let us suppose for contradiction that $\alpha$ is an outer curve in $\Sigma_1$ sent by $\phi$ to a nonouter curve in $\Sigma_2$. We note that $\phi(\alpha)$ can not be a separating curve, for $\alpha$ can never correspond to a cut point in a pants adjacency graph, and so $\phi(\alpha)$ must be a nonseparating curve. If $\kappa(\Sigma_1)$ is at least four then we can extend $\alpha$ to a pants decomposition $P$ in which the two curves adjacent to $\alpha$, denoted $\gamma_1$ and $\gamma_2$, are not adjacent in the adjacency graph of $P - \{\alpha\}$. As $\alpha$ is an outer curve, we note that $\gamma_1$ and $\gamma_2$ are adjacent in $P$. According to Lemma 8, $\phi(\gamma_1)$ and $\phi(\gamma_2)$ can only, together with $\partial \Sigma_2$, border a four-holed sphere containing $\phi(\alpha)$. However, by assumption $\phi(\alpha)$ is not an outer curve. Therefore, $\phi(\gamma_1)$ and $\phi(\gamma_2)$ are not adjacent in $\phi(P)$ and this is contrary to the statement of Lemma 7. (See Figure 2 for one example.)

Once more, the only remaining nontrivial case in low complexity is that of the three-holed torus. Suppose that $\alpha$ is an outer curve sent to a nonseparating curve by $\phi$. Extend $\alpha$ to a pants decomposition $P$ containing a separating curve. Then
the nonseparating curve in $P$ is sent to an outer curve by $\phi$, and this is contrary to Lemma 10.

It now follows that small subsurfaces can not change topological type under embeddings.

**Lemma 12.** Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$ and that when both have equal complexity at most three they are homeomorphic and not the two-holed torus. Let $Z$ be any essential subsurface of $\Sigma_1$ of complexity one and bordered by a single curve $\beta$. Then, for any embedding $\phi$ from $\mathcal{E}(\Sigma_1)$ to $\mathcal{E}(\Sigma_2)$, the essential subsurface $\phi(Z)$ of $\Sigma_2$ filled by $\phi(X(Z))$ is homeomorphic to $Z$.

**Proof.** Such a change in topology would otherwise force $\phi$ to send a nonseparating curve to an outer curve or an outer curve to a nonseparating, contrary to Lemmas 10 and 11, respectively.

We can finally rule out cross-embeddings, and thereafter we regard the two surfaces as being equal and denote both by $\Sigma$.

**Lemma 13.** Suppose that $\Sigma_1$ and $\Sigma_2$ are two compact and orientable surfaces such that the complexity of $\Sigma_1$ is at least that of $\Sigma_2$, and that whenever they have complexity at most two they are homeomorphic and whenever they have complexity equal to three they are either homeomorphic or one is the three-holed torus. Then, there is an embedding $\phi : \mathcal{E}(\Sigma_1) \to \mathcal{E}(\Sigma_2)$ only if $\Sigma_1$ and $\Sigma_2$ are homeomorphic.

**Proof.** The existence of such an embedding implies the complexities $\kappa(\Sigma_1)$ and $\kappa(\Sigma_2)$ are equal. When $\kappa(\Sigma_1)$ is at least four, we know that any such embedding must send separating curves to separating curves. We recall that the size of a maximal collection of distinct and disjoint separating curves in $\Sigma_1$ is precisely $2\text{genus}(\Sigma_1) + |\partial \Sigma_1| - 3$. By our earlier work, this is at most $2\text{genus}(\Sigma_2) + |\partial \Sigma_2| - 3$ and so $\text{genus}(\Sigma_1) \geq \text{genus}(\Sigma_2)$.

To prove equality, we take $Q$ to be a maximal collection of distinct and disjoint curves on $\Sigma_1$ each bounding a one-holed torus. That is, $Q$ has $\text{genus}(\Sigma_1)$ curves. According to Lemma 12, each curve in $\phi(Q)$ must also bound a one-holed torus in $\Sigma_2$. We deduce $\text{genus}(\Sigma_1) \leq \text{genus}(\Sigma_2)$. Combining the two inequalities we have $\text{genus}(\Sigma_1) = \text{genus}(\Sigma_2)$, and $\Sigma_1$ and $\Sigma_2$ are homeomorphic.

Turning to the low complexity surfaces, there are no embeddings from the curve complex of the six-holed sphere or closed surface of genus two to that of the three-holed torus. To see this, extend an outer or nonseparating curve $\alpha$ in $\Sigma_1$ to a pants decomposition $P$ consisting only of outer or nonseparating curves, respectively, and choose a separating curve $\beta$ disjoint from both curves in $P - \{\alpha\}$ and therefore of small intersection with $\alpha$. We may assume that if any curve in $\phi(P)$ is outer
then it is \( \phi(\alpha) \). Now \( \phi(\beta) \) is a nonouter separating curve intersecting \( \phi(\alpha) \) and it follows that \( \phi(\beta) \) must intersect another curve in \( \phi(P) \). This is a contradiction.

The remaining cases, namely from the curve complex of the three-holed torus to the curve complex of the six-holed sphere or of the closed surface of genus two, are covered as follows: For any pants decomposition \( P \) in \( \Sigma_1 \) comprising only of nonseparating curves, choose a nonouter separating curve \( \beta \) meeting only two curves in \( P \). When \( \Sigma_2 \) is the six-holed sphere, according to Lemma 8 each curve in \( P \) can only go to an outer curve. By Lemma 6, small intersection is preserved. Now any nonouter separating curve in the six-holed sphere meets either only one curve or all three curves in a pants decomposition made up entirely of outer curves. It follows that \( \phi(\beta) \) meets every curve in \( \phi(P) \), and this is a contradiction. This simultaneously deals with \( \Sigma_2 \) the closed surface of genus two. \( \square \)

To allow the induction argument to pass through complexity one surfaces unhindered, we need the following lemma on minimal intersection in those subsurfaces bordered by a single curve. This relies on what is a well-established argument, given in [Ivanov 1997] for intersection one and in [Korkmaz 1999; Luo 2000] for intersection two with zero algebraic intersection. Although both are stated for automorphisms, both apply in our setting.

**Lemma 14.** Suppose that \( \Sigma \) is a compact and orientable surface of positive complexity and not homeomorphic to the two-holed torus. Suppose that \( Z \) is an essential subsurface of \( \Sigma \) of complexity one and bordered by a single curve \( \beta \). Then, any embedding \( \phi : \mathcal{E}(\Sigma) \rightarrow \mathcal{E}(\Sigma) \) preserves minimal intersection and its type on \( X(Z) \).

This closes our study of the topological properties of curve complex embeddings, and the promised induction argument now starts with a look at the Farey graph.

**Lemma 15.** Every simplicial embedding from a Farey graph \( \mathcal{F} \) to itself is an automorphism.

**Proof.** We note that each edge in \( \mathcal{F} \) separates and belongs to exactly two 3-cycles and that such a map sends 3-cycles to 3-cycles. Thus, any embedding \( \phi \) on \( \mathcal{F} \) induces an embedding \( \phi^* \) on the dual graph. This graph is a tree in which every vertex has the same finite valence, hence the induced map is a surjection. It follows that every 3-cycle of \( \mathcal{F} \) is contained in the image of \( \phi \). That is to say, \( \phi \) is also a surjection. \( \square \)

It is a well-known fact (indeed, it was known to Dehn [1987, paper 8]) that the automorphisms of \( \mathcal{E}(\Sigma) \) are all induced by surface homeomorphisms when \( \Sigma \) is either a four-holed sphere or a one-holed torus. This completes the base case of the induction.
We now furnish the inductive step. Let \( \phi : \mathcal{C}(\Sigma) \rightarrow \mathcal{C}(\Sigma) \) be any embedding satisfying the hypotheses of Theorem 1. Let \( \alpha \) be any curve in \( \Sigma \). Our previous work on the topological properties of \( \phi \) tells us that the complement of \( \alpha \) and the complement of \( \phi(\alpha) \) are homeomorphic. Therefore, after first composing with a suitable mapping class if need be, \( \phi \) restricts to a self-embedding on the curve complex associated to each component of \( \Sigma - \alpha \). The embeddings arising in this way are very natural for they inherit many of the properties verified by \( \phi \), for instance they also preserve the separating type of a curve. This is of particular relevance when cutting the surface \( \Sigma \) along a curve and finding a two-holed torus complementary component. In [Luo 2000], the author explains how to find automorphisms of the curve complex associated to the two-holed torus not induced by a surface homeomorphism. No such automorphism can arise as a restriction, nor can any embedding, as outer curves in this two-holed torus correspond to separating curves in \( \Sigma \).

Our inductive hypothesis therefore applies and it tells us that each restriction of \( \phi \) associated to a positive complexity component of \( \Sigma - \alpha \) is induced by a surface homeomorphism. In gluing back together by identifying the boundary components of \( \Sigma - \alpha \) corresponding to \( \alpha \), we have a countable family of mapping classes where each such mapping class \( f \) satisfies \( f(\delta) = \phi(\delta) \) for all \( \delta \in X(\alpha) \cup \{\alpha\} \). We must somehow decide which of these, if any, is appropriate.

This construction applies equally well for every curve on \( \Sigma \), in particular any curve \( \beta \) adjacent to \( \alpha \). The set of mapping classes associated to \( \alpha \) and the set of mapping classes associated to \( \beta \) have nonempty intersection. That is, to the edge of \( \mathcal{C}(\Sigma) \) spanned by \( \alpha \) and \( \beta \), we can associate at least one mapping class \( f \) with \( f(\delta) = \phi(\delta) \) for all \( \delta \in X(\alpha) \cup X(\beta) \).

We need to verify that for any three curves \( \alpha, \beta_1, \beta_2 \) such that \( \alpha \) is adjacent to both \( \beta_1 \) and \( \beta_2 \), the action on \( \mathcal{C}(\Sigma) \) of any such mapping class \( f_1 \) associated to the edge \( \alpha, \beta_1 \) agrees with that of any such mapping class \( f_2 \) associated to the edge \( \alpha, \beta_2 \). For almost all surfaces, \( f_1 \) and \( f_2 \) will be the same mapping class. For now there remains the possibility that \( \Sigma - \alpha \) has an exceptional surface component and in particular the possibility that \( f_1^{-1} f_2 \) Dehn twists around \( \alpha \), or a combination of the two. We treat this in the following lemma.

**Lemma 16.** Suppose that \( \alpha, \beta_1, \beta_2 \in X(\Sigma) \) are distinct, with \( \beta_1 \) and \( \beta_2 \) of zero or otherwise minimal intersection and \( \alpha \) disjoint from both \( \beta_1 \) and \( \beta_2 \). Suppose \( f_1, f_2 \in \text{Map} \Sigma \) are two mapping classes such that \( f_i(\delta) = \phi(\delta) \) for all \( \delta \in X(\alpha) \cup X(\beta_i) \), for \( i = 1, 2 \). Then, \( f_1^{-1} f_2 \in \text{Ker} \Sigma \).

**Proof.** Let \( f \) denote the mapping class \( f_1^{-1} f_2 \), noting \( f \) acts trivially on \( X(\alpha) \), and suppose for contradiction that \( f \notin \text{Ker} \Sigma \). As we shall see in the subsequent paragraphs, there then exist disjoint (possibly equal) curves \( \delta_1 \) and \( \delta_2 \) on \( \Sigma \).
Figure 3. The case of $\Sigma$ a five-holed sphere; $\delta_1$ is the only curve on $\Sigma$ disjoint from both $\beta_1$, $\delta_2 \in X(\delta_1) \cap (X(\alpha) \cup X(\beta_2))$. As it happens, $X(\delta_1) \cap (X(\alpha) \cup X(\beta_2))$ equals $\{\beta_1, \delta_2\}$ in this instance.

such that at least one of $\iota(\delta_i, \alpha)$ and $\iota(\delta_i, \beta_i)$ is zero, for both $i = 1, 2$, and such that $\iota(\delta_1, f(\delta_2)) > 0$. Given this, we also have $\iota(\delta_1, f(\delta_2)) = \iota(\delta_1, f^{-1}_1 f^{-1}_2(\delta_2)) = \iota(f_1(\delta_1), f_2(\delta_2)) = \iota(\phi(\delta_1), \phi(\delta_2)) = 0$. This is a contradiction, and we deduce the statement of the lemma.

To see that such a pair of curves $\delta_1, \delta_2$ must exist, we can argue as follows. Suppose $\delta_1 \in X(\beta_1)$ has minimal intersection with $\alpha$ and zero or minimal intersection with $\beta_2$. Then, $\delta_1$ is entirely determined by the nonempty set $X(\delta_1) \cap (X(\alpha) \cup X(\beta_2))$. More precisely, $\delta_1$ is the only curve on $\Sigma$ intersecting $\alpha$ and disjoint from every curve in $X(\delta_1) \cap (X(\alpha) \cup X(\beta_2))$. See Figure 3 for one example in the five-holed sphere, in this case a pentagon configuration as described in [Luo 2000].

Suppose for contradiction that $\iota(\delta_1, f(\delta_2)) = 0$ for any curve $\delta_2 \in X(\delta_1) \cap (X(\alpha) \cup X(\beta_2))$. Then, $f(X(\delta_1) \cap (X(\alpha) \cup X(\beta_2))) \subseteq X(\delta_1) \cap (X(\alpha) \cup X(\beta_2))$. However, because $f$ is a mapping class this inclusion is an equality and we deduce $f(\delta_1) = \delta_1$. As the complement in $\Sigma$ of $\beta_1$ is filled by a set of curves all fixed by $f$, we deduce $f$ acts trivially on $X(\beta_1)$. Arguing along similar lines, by reinterpreting our contention as $\iota(\delta_2, f^{-1}(\delta_1)) = 0$ we deduce $f$ acts trivially on $X(\beta_2)$ as well.

We have shown that $f(\delta) = \delta$ for all $\delta \in X(\alpha) \cup X(\beta_1) \cup X(\beta_2)$. However $X(\alpha) \cup X(\beta_1) \cup X(\beta_2)$ fills $\Sigma$, that is every curve on $\Sigma$ has nonzero intersection with some curve from this set. It follows that $f$ in fact fixes every curve on $\Sigma$. Therefore, $f \in \text{Ker} \Sigma$ and by assumption this is absurd. □
The link of $\alpha$ is either chain-connected, so that for any two of its vertices, $\beta_1$ and $\beta_2$, there is a sequence of curves $\beta_1 = \delta_1, \delta_2, \ldots, \delta_n = \beta_2$ each distinct and disjoint from $\alpha$ and such that consecutive curves $\delta_i, \delta_{i+1}$ have minimal intersection, or is connected. By applying Lemma 16 inductively, we conclude that any two edges ending on $\alpha$ are prescribed the same automorphism of $\mathcal{C}(\Sigma)$ and that any such automorphism is induced by a mapping class. Since $\mathcal{C}(\Sigma)$ is connected, it follows that every edge is allocated the same such automorphism $\Phi$.

All we need do now is verify that this automorphism is equal to $\phi$ everywhere. To do this, we only need to remark that any curve $\alpha$ spans an edge with a second curve $\beta$. This edge is prescribed the automorphism $\Phi$ which, by construction, agrees with $\phi$ on both $X(\alpha)$ and $X(\beta)$. In particular, $\Phi$ agrees with $\phi$ on $X(\beta)$ which contains $\alpha$. This completes one proof of Theorem 1.

Acknowledgements

The author thanks the Department of Mathematical and Computing Sciences at the Tokyo Institute of Technology for its warm and generous hospitality, and it is his pleasure to thank Brian Bowditch, Javier Aramayona, Jason Behrstock, Dan Margalit, Graham Niblo and Caroline Series for their enthusiasm and for being so generous with their time; their comments on this work improved the exposition beyond his own capabilities. Last, but by no means least, he wishes to thank the EPSRC for supporting this research.

References


Received July 27, 2005.

KENNETH J. SHACKLETON
DEPARTMENT OF MATHEMATICAL AND COMPUTING SCIENCES
TOKYO INSTITUTE OF TECHNOLOGY
2-12-1 O-OYAYAMA MEGURO-KU
TOKYO 152-8552
JAPAN
shackleton.k.aa@m.titech.ac.jp
kjs2006@alumni.soton.ac.uk
A GIAMBELLI-TYPE FORMULA
FOR SUBBUNDLES OF THE TANGENT BUNDLE

Boris Shapiro and Maxim Kazarian

Consider a generic $n$-dimensional subbundle $\mathcal{V}$ of the tangent bundle $TM$ on some manifold $M$. Given $\mathcal{V}$, one can define different degeneracy loci $\Sigma_r(\mathcal{V})$, $r = (r_1 \leq r_2 \leq \cdots \leq r_k)$, on $M$ consisting of all points $x \in M$ for which the subspaces $\mathcal{V}^j(x) \subset TM(x)$ spanned by all length $\leq j$ commutators of vector fields tangent to $\mathcal{V}$ at $x$ has dimension less than or equal to $r_j$. Under a certain transversality assumption, we explicitly calculate the $\mathbb{Z}_2$-cohomology classes of $M$ dual to $\Sigma_r(\mathcal{V})$, using determinantal formulas due to W. Fulton and the expression of the Chern classes of the associated bundle of the free Lie algebras in terms of the Chern classes of $\mathcal{V}$.

1. Preliminaries and results

History and motivation. The question of the existence of a nontrivial subbundle of the tangent bundle on a given manifold is a geometric problem of long-standing interest. (Such subbundles are often called distributions, and we will freely use both terms below.) In the basic nontrivial case of rank-2 subbundles, the first important results in the area go back to the classical treatise [Hirzebruch and Hopf 1958]. Apparently, the best achievements in this problem were obtained in late 1960s by Thomas [1967a; 1967b]; see also the well-written survey [Thomas 1969]. Later, some of his results were rediscovered by Matsushita [1988]. Not much has been done in this area since then. One of the few recent exceptions is [Jacobowitz and Mendoza 2003]. Rather detailed information is available about the existence of (oriented) subbundles of rank 2. For rank 3 and higher, only the first obstruction to the problem is known, see [Thomas 1967a]. The algebraic invariants for this result come from the Stiefel–Whitney classes of elements of $K\tilde{O}(M)$, which is the reduced real $K$-theory group of the manifold. Starting with the late 1970s, the interest in the geometric properties of subbundles of the tangent bundle were stimulated by the development of singularity theory and the revival of interest in nonholonomic mechanics. A nice source of information about this topic is [Montgomery 2002]. In particular, if a given manifold admits a subbundle of rank at least


Keywords: $n$-subbundles, free Lie algebra, determinantal formulas.
2, one can construct at each point of the manifold an incomplete flag whose $i$-th subspace is the linear span of the commutators of length at most $i$ of the vector fields tangent to the subbundle. The ranks of these subspaces will (in general) depend on the point; see below.

For a small generic perturbation of the original subbundle, the ranks of the subspaces of the incomplete flags will stay constant almost everywhere on the manifold, and depend only on the rank of the original subbundle and the dimension of the manifold. Generalizing the question with which we started, one can formulate this problem:

**Problem.** When does a manifold admit a distribution whose associated flags have constant (and maximal-possible) ranks throughout the manifold?

Being in general even more difficult than the (still unsolved) initial question, the latter problem has a nice answer in the case of oriented rank-2 distributions on oriented 4-manifolds. A rank-2 distribution on a 4-manifold whose associated flag has the set of ranks $(2, 3, 4)$ at each point is called an Engel distribution. These distributions have remarkable properties; see example [Gershkovich 1995a; 1995b]. For instance, combining the results of this second paper (see also [Kazarian et al. 1997]) with the recent [Vogel 2004; 2005], we get:

**Theorem.** An orientable 4-manifold admits an orientable Engel distribution if and only if it is parallelizable.

The aim of this paper is to develop the basics of the obstruction theory suited to the problem above.

**Standard definitions and notation.** The following notions are standard, see for example [Gershkovich and Vershik 1988]. Let $M$ be an $m$-dimensional manifold and $\mathcal{V} \subset TM$ an $n$-dimensional subbundle (a rank $n$-distribution) on $TM$. Given $\mathcal{V}$, one associates at each point $x \in M$ its derived flag

\[ f\mathcal{V}_x = \{ \mathcal{V}_x^1 \subseteq \mathcal{V}_x^2 \subseteq \cdots \subseteq \mathcal{V}_x^j \subseteq \cdots \}, \]

where $\mathcal{V}_x^j = \{ \mathcal{V}_x^{j-1} + [\mathcal{V}_x^{j-1}, \mathcal{V}_x] \}$. If at each point $x \in M$ there exists a positive integer $k(x)$ such that the subspace $\mathcal{V}_x^{k(x)}$ coincides with $TM_x$, then $\mathcal{V}$ is called (maximally) nonholonomic. Let $n_j(x)$ denote the dimension of $\mathcal{V}_x^j$. The set of numbers $(n_1(x), \ldots, n_k(x))$ is called the growth vector of $\mathcal{V}$ at $x$. For a given nonholonomic $\mathcal{V}$, the minimal number $k$ such that $\mathcal{V}_x^k = TM_x$ at all points $x$ is called the degree of nonholonomicity. (A subbundle $\mathcal{V}$ is called regular if the $n_j(x)$ do not depend on $x$; the corresponding set of numbers $(n_1, n_2, \ldots, n_k)$ is called the growth vector of the regular subbundle $\mathcal{V}$.) Let $\mathfrak{L}_n$ denote the free Lie algebra with $n$ generators, and let $\mathfrak{L}_n^k$ be its linear subspace spanned by all elements of length $k$. Let $d(n, k)$ be the dimension of $\mathfrak{L}_n^k$. An $n$-dimensional $\mathcal{V}$ with degree of
nonholonomicity $k$ is called a maximal-growth subbundle (or an MG-distribution) if $n_k = m$ and $n_j = \sum_{i \leq j} d(n, j)$ for all $j < k$. The growth vector with the entries 

$$(n = d(n, 1), d(n, 1) + d(n, 2), \ldots, d(n, 1) + d(n, 2) + \ldots + d(n, k), m)$$

is called the maximal growth vector.

**Remark.** According to [Gershkovich 1988], a germ of a generic distribution has maximal growth. Here, generic means belonging to some open everywhere-dense subset in the $C^\infty$-Whitney topology. Thus, locally a typical subbundle is an MG-distribution, while globally there are (many) topological obstructions to the existence of MG-distributions on a given $M$. The problem we are addressing in the present paper can be reformulated as constructing obstructions to the existence of MG-distributions on a given manifold.

**Examples.** A contact structure is a regular MG-distribution. A 2-dimensional MG-subbundle on a four-dimensional manifold is called an Engel distribution; see above and [Gershkovich 1995b]. This is the only example of distributions with a stable local normal form, besides contact structures and their even-dimensional analogs.

**Degeneracy loci.** Given a generic $n$-distribution $\mathcal{V} \subset TM$, one expects that globally $M$ contains a (typically, reducible) degeneracy locus $\Sigma$ consisting of all such points $x$ where the growth vector $(n_1(x), \ldots, n_k(x))$ is lexicographically smaller than the maximal one. Given a growth vector $r = (r_1 \leq r_2 \leq \cdots \leq r_k)$, denote by $\Sigma_r$ the subset of all $x \in M$ satisfying the conditions $n_1(x) \leq r_1, \ldots, n_k(x) \leq r_k$. Such $\Sigma_r$ can be considered as degeneracy loci in the standard meaning of intersection theory; see Section 2 (compare with [Fulton 1984, Chapter 14]). Namely, each $n$-distribution $\mathcal{V} \subset TM$ induces the associated fiber bundle $\mathcal{L}(\mathcal{V}) \to M$, where the fiber $\mathcal{L}_x$ is the free Lie algebra generated by the subspace $\mathcal{V}_x$. The bundle $\mathcal{L}(\mathcal{V})$ has an obvious grading $\mathcal{L}(\mathcal{V}) = \bigoplus_{k=1}^\infty \mathcal{L}^k(\mathcal{V})$ coming from $\mathcal{L}_n = \bigoplus_{k=1}^\infty \mathcal{L}_n^k$. Moreover, one can define a natural map $\Phi : \mathcal{L}(\mathcal{V}) \to TM$ of vector bundles, sending each $\bigoplus_{i=1}^k \mathcal{L}^k(\mathcal{V})$ onto $\mathcal{V}^k$. (The map $\Phi$ is not unique, in very much the same way as the identifying map between $TM$ and its nilpotentization; compare with [Gershkovich and Vershik 1988].) This allows us to apply the determinantal formulas of [Fulton 1992] to the map $\Phi$, under the assumption that the considered $\Sigma_r$ has the expected (co)dimension, that is, the same codimension as the corresponding degeneracy locus for a generic map of flag bundles of the dimensions prescribed by $\mathcal{L}_n$. Algebraically, in order to be able to apply these formulas one also needs to express the Chern classes of $\mathcal{L}(\mathcal{V})$ in terms of those of $\mathcal{V}$.

**Remark.** Fulton [1992] has generalized a large number of previously known special cases of determinantal formulas giving the cohomology classes of different
degeneracy loci for the maps of vector bundles to a very general situation of maps of flagged vector bundles. Such formulas could be traced back (through the works of Porteous, Thom, Laksov and Kempf, and many other authors) to the pioneering results of Giovanni Giambelli on the degrees of different strata in the spaces of matrices. For this reason in a number of publications similar determinantal formulas are named after Giambelli; see, for example, [Fulton 1992, §7]. For a detailed account on determinantal formulas and degeneracy loci we recommend [Fulton and Pragacz 1998]; for information on Giambelli, see [Laksov 1994].

The contents of the paper are as follows. In Section 2 we construct the map $\Phi : \mathfrak{L}(\mathcal{V}) \to TM$. In Section 3 we find an explicit formula for the Chern character of the bundle $\mathfrak{L}^k(\mathcal{V})$ in terms of the Chern character of $\mathcal{V}$; it turns out to be similar to the formula for the dimensions of $\Omega_n^k$. In principle, this allows us to calculate the Chern classes of $\text{Lie}^k(\mathcal{V})$ for any reasonable specific example by inverting Newton polynomials; see the Appendix. In Section 4 we recall the appropriate determinantal formula for maps of flag bundles, adjust it to our needs and calculate some examples. Section 5 is devoted to some generalities on derived flags $\mathcal{F}\mathcal{V}$ and the standard stratification of the spaces of matrices as well as counterexamples to transversality in big codimensions. In Section 6 we enumerate all potentially admissible defect vectors occurring for a generic distribution and prove the necessary transversality result showing that $\Sigma_r$ have the expected (co)dimension in the cases $n = 2$, $m \leq 8$ and $n \geq 3$, $m \leq \frac{1}{6}n(n+1)(2n+1)$. In Section 7 we briefly discuss some further directions of study and possible generalizations of the transversality theorem. The main result of the paper is formula (11) justified by the transversality theorem for the above mentioned values of $(n, m)$. Finally, appendix contains the Mathematica code which explicitly calculates the necessary Chern classes of the homogeneous components of the free Lie algebra $\mathfrak{L}_n^k$ up to order 4.

In short, the main proposition of the paper can be summarized as follows. In order to calculate the universal formula for the cohomology class dual to $\Sigma_r$ (or rather for the Stiefel–Whitney) classes of $\bigoplus_{i=1}^k \mathfrak{L}^i(\mathcal{V})$ and $TM$ into the appropriate Giambelli-type formula.

2. The $\Sigma_r$ as classical degeneracy loci

Associated bundle of free Lie algebras. Given $\mathcal{V} \subset TM$, we define the map $\Phi : \mathfrak{L}(\mathcal{V}) \to TM$ such that each subbundle $\bigoplus_{i=1}^k \mathfrak{L}^i(\mathcal{V})$ is mapped onto $\mathcal{V}^j$. In fact we define another filtered vector bundle $\mathfrak{N}(\mathcal{V})$ whose associated graded bundle is isomorphic to $\mathfrak{L}(\mathcal{V})$ with the canonical map $\Psi : \mathfrak{N}(\mathcal{V}) \to TM$. As a result one gets a (nonunique) map $\Phi : \mathfrak{L}(\mathcal{V}) \to TM$ defined up to a filtered isomorphism between $\mathfrak{N}(\mathcal{V})$ and $\mathfrak{L}(\mathcal{V})$. 
Universal map. Let $\mathcal{V} \subset TM$ be an $n$-subbundle and $k$ be its degree of nonholonomicity.

**Theorem 1.** There exists a globally defined map of vector bundles

$$\Phi_k : \bigoplus_{i \leq k} \mathcal{L}^i(\mathcal{V}) \to TM$$

such that for all $x \in M$ and $j \leq k$, the subspace $\mathcal{V}^j(x)$ coincides with the image of $\Phi_j(x) : \bigoplus_{i \leq j} \mathcal{L}^i(\mathcal{V})(x) \to T_x M$, where $\Phi_j$ is the restriction of $\Phi_k$ to $\bigoplus_{i \leq j} \mathcal{L}^i(\mathcal{V}) \subset \bigoplus_{i \leq k} \mathcal{L}^i(\mathcal{V})$.

For the local version of this theorem, see [Gershkovich and Vershik 1988].

**Proof.** We construct an auxiliary flag of vector bundles

$$N^1(\mathcal{V}) \subset N^2(\mathcal{V}) \subset \ldots \subset N^k(\mathcal{V}),$$

the map $\Psi_k : N^k(\mathcal{V}) \to TM$ satisfying the statement of the Theorem, and a canonical isomorphism

$$N^j(\mathcal{V})/N^{j-1}(\mathcal{V}) \cong \mathcal{L}^j(\mathcal{V})$$

which shows that the flag of bundles $N(\mathcal{V})$ is isomorphic to the flag of bundles $\mathcal{L}(\mathcal{V})$.

Recall the notion of a Hall basis in the free Lie algebra $\mathcal{L}_n$ with $n$ generators, see [Bourbaki 1975]. Namely, $\mathcal{L}_n$ has the following standard graded basis $H$ called Hall family or Hall basis. Given a linearly ordered set $V$ (of cardinality $n$), we define the following linearly ordered subset $H$ in the free monoid $\text{Mon}_V$.

1. If $u, v \in H$ and $\text{lng}(u) < \text{lng}(v)$ then $u < v$ where $\text{lng}$ denotes the usual length of a word in $\text{Mon}_V$.
2. $V = H^1 \subset H$ and $H^2$ consists of the set of all ordered pairs $(v_1, v_2) \in H$ where $v_1 < v_2$.
3. Each element of $H$ of length at least 3 has the form $a(bc)$, where $a, b, c \in H$, $bc \in H$, $b \leq a < bc$ and $b < c$. (Obviously, $H = \bigcup_{k=1}^\infty H^k$, where $H^k$ is the set of all length $k$ elements in $H$.)

**Examples.** Suppose $V = \{u, v\}$. Then

- $H^1 = \{u, v\}, \quad H^2 = \{(u, v), (v(u, v))\}$,
- $H^3 = \{(u(u, v)), (v(v(u, v)))\}$,
- $H^4 = \{(u(u(u, v))), (v(v(v(u, v))))\}$,
- $H^5 = \{(u(u(u(u, v)))), (v(v(v(v(u, v)))))\}$. 


The construction of the flag of bundles \( N(\mathcal{V}) \) is as follows. Let \( W \) be the sheaf of free Lie algebras associated to the sheaf of local sections of the bundle \( \mathcal{V} \subset TM \). The elements of \( W \) are \( \mathbb{R} \)-linear combinations of Lie monomials of sections of \( \mathcal{V} \). Denote by \( \mathcal{A} \) the sheaf of rings of smooth functions on \( M \). Define the homomorphism \( D : W \to \text{der} \mathcal{A} \) as follows. If \( v \in \mathcal{V} \subset W \) is of degree 1 we put

\[
D_v f = vf,
\]

i.e., the usual Lie derivative of the function \( f \) along the vector field \( v \). Then, we assign by induction

\[
D_{[a,b]} f = D_b(D_a f) - D_a(D_b f).
\]

The operation \( D \) is well defined and \( D_u \) is the derivation of \( \mathcal{A} \) for every \( u \in W \). Consider the sheaf \( \mathcal{A} \otimes \mathcal{W} \) of \( \mathcal{A} \)-modules. We introduce the Lie algebra structure on \( \mathcal{A} \otimes \mathcal{W} \) as

\[
[f \otimes u, g \otimes w] = fg \otimes [u, w] + f D_u g \otimes w - g D_w f \otimes u
\]

(the Jacobi identity is verified by direct computation). Finally, define \( N(\mathcal{V}) \) to be the quotient Lie algebra of \( \mathcal{A} \otimes \mathcal{W} \) by the ideal generated by all relations of the form

\[
f \otimes v = 1 \otimes fv, \quad f \in \mathcal{A}, \quad v \in \mathcal{V} \subset W.
\]

Having (4) in mind we drop the sign of the tensor product in the notation of the elements in \( N(\mathcal{V}) \). The filtration on \( W \) by the length of Lie monomials gives the natural filtration \( N^j(\mathcal{V}) \) on the sheaf \( N(\mathcal{V}) \) of \( \mathcal{A} \)-modules. We claim that all \( N^j(\mathcal{V}) \) are locally free sheaves of \( \mathcal{A} \)-modules of finite ranks. Indeed, let \( e_1, \ldots, e_n \) be the set of local sections of \( \mathcal{V} \) over some open domain \( U \subset M \) such that these sections form a basis in each fiber \( \mathcal{V}(x), x \in U \). Then it follows from (1)–(4) that every section \( u \) of \( W \) over \( U \) can be represented as

\[
u = \sum f_i h_{i}(e_1, \ldots, e_n),
\]

where \( f_i \) are some functions and \( h_{i} \in \mathcal{H} \) are the elements of Hall basis of the free Lie algebra \( \mathfrak{L}_n \). Moreover, this representation is unique, i.e. the set of sections \( h_{i}(e_1, \ldots, e_n), l \leq \dim \bigoplus_{i \leq j} \mathfrak{L}_n^i \) forms the set of free generators of the \( \mathcal{A} \)-module \( N^j(\mathcal{V}) \). Thus the \( \mathcal{A} \)-module \( N^j(\mathcal{V}) \) is the module of sections of some vector bundle which we also denote as \( N^j(\mathcal{V}) \).

Observe that if \( [u, w] \in N(\mathcal{V}) \) has degree \( j \) then \( [fu, gw] - fg[u, w] \) has degree strictly less than \( j \). Therefore, the homomorphism \( N^j(\mathcal{V})/N^{j-1}(\mathcal{V}) \to \mathfrak{L}^j(\mathcal{V}) \) is well defined. Moreover, the arguments above show that this homomorphism is, in fact, an isomorphism of vector bundles.
The homomorphism of Lie algebras $\Psi : N(\mathcal{V}) \rightarrow \text{Vect}(M)$ is now obvious. It sends a formal Lie bracket of vector fields in $\mathcal{V}$ to the corresponding commutator of these vector fields. Formulas (1)–(4) show that this homomorphism is well defined and $\Psi(N^j(\mathcal{V})(x))$ coincides with $\mathcal{V}^j(x)$ by definition.

Remark. The vector bundle $N^j(\mathcal{V})$ can be also described, as a usual vector bundle by trivializations and transition functions. Trivializations of $N^j(\mathcal{V})$ correspond to the trivializations of $\mathcal{V}$ and are given by the sections $h_t(e_1, \ldots, e_n)$, $l \leq \dim \bigoplus_{i \leq j} \mathcal{L}_n^i$, where $e_1, \ldots, e_n$ are sections giving some local basis of $\mathcal{V}$. If $\{e'_1, \ldots, e'_n\}$ is another basis such that $e_i = \sum a_ir_i e'_j$ then to find transition functions for $N^j(\mathcal{V})$ one should express $h_t(\sum a_1 r p e'_p, \ldots, \sum a_n r p e'_p)$ using (1)–(4) as a linear combination of $h_t(e'_1, \ldots, e'_n)$ with some functional coefficients.

3. On the Chern classes of the bundle of free Lie algebras

Let $E \to M$ be a complex vector bundle of dimension $n$ over a smooth compact manifold $M$ (not necessarily a subbundle of $TM$). For any linear representation of the group $\text{GL}(n, \mathbb{C})$ in $\mathbb{C}^n$ one can associate in a natural way to the bundle $E \to M$ the corresponding $m$-dimensional bundle over $M$. For example, the bundles $\text{End}(E)$, $\Lambda^2 E$ etc. are associated with the obvious representations of $\text{GL}(n, \mathbb{C})$ in $\mathbb{C}^n \otimes \mathbb{C}^n$, $\mathbb{C}^n$, $\Lambda^2 \mathbb{C}^n$ respectively.

Given a basis $\{e, \ldots, e_n\}$ in $\mathbb{C}^n$ let $\mathfrak{L}_n$ denote the free Lie algebra with the generators $\{e, \ldots, e_n\}$, and let $\mathfrak{L}_n^k$ be its $k$-th homogeneous component. A linear change of the above basis acts naturally on the spaces $\mathfrak{L}_n^k$. Denote by $\mathfrak{L}_n^k(E)$ or simply by $\mathfrak{L}_n^k$ the bundle over $M$ associated with this action. The relation between the characteristic classes of the bundles $E$ and $\mathfrak{L}_n^k(E)$ is described in the following theorem. (This question was already proposed in [Thrall 1942].)

Let $\text{ch}(E) \in H^*(M)$ be the Chern character of a bundle $E$. For any element $\eta = \eta_0 + \eta_1 + \eta_2 + \cdots \in H^*(M)$ and a number $d$ set $(\eta)_d = \eta_0 + \eta_1 d + \eta_2 d^2 + \cdots$, where $\eta_i$ belongs to $H^{2i}(M)$.

**Theorem 2** (see [Reutenauer 1993], for example). **The Chern character of the bundle $\mathfrak{L}_n^k$ is given by**

\[
(5) \quad \text{ch}(\mathfrak{L}_n^k) = \frac{1}{k} \sum_{d|k} \mu(d) \left(\text{ch}(E)^{k/d}\right)_d.
\]

Here the summation is taken over the set of all divisors of $k$ and $\mu$ is the Möbius function. By taking the component of degree 0 in this formula we get the well-known expression

\[
\text{dim}(\mathfrak{L}_n^k) = \frac{1}{k} \sum_{d|k} \mu(d) n^{k/d}.
\]
for the dimension of $\mathfrak{L}^k_n$ [Bourbaki 1975].

Proof. The main observation is that the total tensor algebra of $\mathbb{C}^n$ is isomorphic, as an $GL(n, \mathbb{C})$-module, to the universal enveloping algebra of $\mathfrak{L}_n$. Therefore, by the Poincaré–Birkhoff–Witt theorem,

$$T^*(E) \cong S^*(\mathfrak{L}(E)) = S^*(\mathfrak{L}^1) \otimes S^*(\mathfrak{L}^2) \otimes \cdots.$$ 

Applying the Chern character to both sides we get

$$\frac{1}{1 - \text{ch}(E)t} = \prod_{n=1}^{\infty} s(\mathfrak{L}^k)(t^k),$$

where $t$ is a formal parameter, $s_i(V) = \text{ch}(S^i V)$ is the Chern character of the $i$-th symmetric power of $V$, and $s(V)$ is a formal series $s(V)(t) = \sum s_i(V)t^i$. Now, applying $-td\log$ to both sides we get

$$\frac{\text{ch}(E)t}{1 - \text{ch}(E)t} = -\sum_{n=1}^{\infty} kt^n(d \log s(\mathfrak{L}^k))(t^k).$$

Observe now that $-d \log s(V)(t) = (\text{ch}(V))_1 + (\text{ch}(V))_2t + (\text{ch}(V))_3t^2 + \cdots$. To prove this we can write (using the splitting principle) $V = V_1 \oplus \cdots \oplus V_m$, where the 1-dimensional bundle $V_i$ has Chern character $h_i$. Then

$$-d \log s(V) = -d \log (s(V_1)s(V_2)\cdots s(V_m)) = -d \log \prod \frac{1}{1 - h_it} = \sum \frac{h_i}{1 - h_it}$$

$$= \sum h_1 + \sum h_i^2 t + \sum h_i^3 t^2 + \cdots$$

$$= (\text{ch}(V))_1 + (\text{ch}(V))_2t + (\text{ch}(V))_3t^2 + \cdots,$$

since $h_i^k = e^{ht} = \sum t^{k/l} / l!$, where $t_i$ is the first Chern class of $V_i$. Therefore, (6) is equivalent to

$$\frac{\text{ch}(E)t}{1 - \text{ch}(E)t} = \sum (\text{ch}(\mathfrak{L}^1))_kt^k + 2 \sum (\text{ch}(\mathfrak{L}^2))_kt^{2k} + 3 \sum (\text{ch}(\mathfrak{L}^3))_kt^{3k} + \cdots.$$ 

Comparing the terms of the same degree in $t$ we get

$$\text{ch}(E)^k = \sum_{d|k} d(\text{ch}(\mathfrak{L}^d))_{k/d}.$$ 

If we now multiply the $l$-th homogeneous component of this equality by $k^{-l}$ then after this rescaling we get

$$(\text{ch}(E)^k)_{1/k} = \sum_{d|k} d(\text{ch}(\mathfrak{L}^d))_{1/d}.$$
Applying to the latter equality the Möbius inversion formula we obtain

\[ k(\text{ch}(L^k))_{1/k} = \sum_{d|k} \mu(d)(\text{ch}(E)^{k/d})_{d/k}, \]

which (after another rescaling) gives the required formula. \( \square \)

**Examples.** The relation between the Chern classes and the Chern character gives the possibility to compute the Chern classes of \( \mathcal{O}_n^k \). For \( k \leq 4 \) (taking in account only terms of degree at most 4 in the characteristic classes) we obtain the following explicit formulas for the total Chern class of \( \mathcal{O}_n^k \):

\[ c(\mathcal{O}_n^1) = c(E) = 1 + c_1 + c_2 + c_3 + c_4 + \cdots, \]

\[ c(\mathcal{O}_n^2) = 1 + (-1 + n)c_1 + ((1 - \frac{3n}{2} + \frac{n^2}{2})c_1^2 + (-2 + n)c_2 \]
\[ + ((-1 + \frac{11n}{6} - n^2 + \frac{1}{6}n^3)c_1^3 + (4 - 4n + n^2)c_1c_2 + (-4 + n)c_3) \]
\[ + ((1 - \frac{25n}{12} n + \frac{5n^2}{2} - \frac{n^3}{2} + \frac{1}{24}n^4)c_1^4 + (-6 + 8n - \frac{7n^2}{2} + \frac{n^3}{2})c_1^2c_2 \]
\[ + (3 - \frac{5n}{2} + \frac{n^2}{4})c_2^2 + (9 - 6n + n^2)c_1c_3 + (-8 + n)c_4) + \cdots, \]

\[ c(\mathcal{O}_n^3) = 1 + (-1 + n^2)c_1 + ((2 - n - \frac{3n^2}{2} + \frac{n^4}{2})c_1^2 + (-3 + n^2)c_2 \]
\[ + ((-4 + 3n + \frac{13n^2 - n^3 - n^4 + \frac{1}{6}n^6})c_1^3 + (12 - 4n - 5n^2 + n^4)c_1c_2 + (-9 + n^2)c_3) \]
\[ + ((8 - \frac{15n}{2} - \frac{61n^2}{12} + \frac{7n^3}{2} + \frac{47}{24}n^4 - \frac{1}{2}n^5 - \frac{5}{24}n^6 + \frac{1}{24}n^8)c_1^4 \]
\[ + (-36 + 19n + \frac{35}{2}n^2 - 5n^3 - 4n^4 + \frac{1}{2}n^6)c_1^2c_2 \]
\[ + (18 - 6n - \frac{7}{2}n^2 + \frac{1}{2}n^4)c_2^2 + (36 - 6n - 11n^2 + n^4)c_1c_3 + (-27 + n^2)c_4) + \cdots, \]

\[ c(\mathcal{O}_n^4) = 1 + (-n + n^3)c_1 + ((1 + n - n^2 - \frac{1}{2}n^3 - n^4 + \frac{1}{2}n^6)c_1^2 + (-2(n + n^3)c_2 \]
\[ + ((-4 - \frac{1}{2}n + 2n^2 + \frac{8}{3}n^3 + \frac{3}{2}n^4 - n^5 - \frac{1}{2}n^6 - \frac{1}{2}n^7 + \frac{1}{6}n^9)c_1^3 \]
\[ + (8 + 4n - 4n^2 - n^3 - 3n^4 + n^6)c_1c_2 + (-4n + n^3)c_3) \]
\[ + ((13 - 2n - \frac{77}{12}n^2 - \frac{35n^3}{4} - \frac{3}{4}n^4 + \frac{5}{2}n^5 + \frac{35}{24}n^6 + n^7 - \frac{1}{2}n^8 - \frac{1}{4}n^9 - \frac{1}{6}n^{10} + \frac{1}{24}n^{12})c_1^4 \]
\[ + (-48 + 12n^2 + 18n^3 + 7n^4 - 5n^5 - \frac{3n^6}{2} - 2n^7 + \frac{1}{2}n^9)c_1^2c_2 \]
\[ + (24 + 4n - 7n^2 - \frac{1}{2}n^3 - 2n^4 + \frac{1}{2}n^6)c_2^2 \]
\[ + (24 + 8n - 5n^2 - n^3 - 5n^4 + n^6)c_1c_3 + (-8n + n^3)c_4) + \cdots. \]

**Remark.** Substituting \( w_i \) for \( c_i \) in these formulas and reducing coefficients mod 2 one gets the expression for the total Stiefel–Whitney class of \( \mathcal{O}_n^k \) in the case of a real \( n \)-dimensional bundle \( E \). Note that the coefficients of the polynomials above have integer values for any \( n \) and therefore their values mod 2 are well defined.
Determinantal formula  and its application

**Determinantal formula.** First we recall a certain formula borrowed from [Fulton 1992]. Assume that we have a flag $A_1 \subset A_2 \subset \cdots \subset A_l$ of the complex vector bundles over a manifold $M$, with ranks $a_1 \leq a_2 \leq \cdots \leq a_l$, and a map

$$h : A_1 \subset A_2 \subset \cdots \subset A_l \to B$$

to a manifold $B$ of dimension $b$. Assume furthermore that the set of nonnegative integers $\kappa_1, \ldots, \kappa_l$ satisfies the inequalities

$$(7) \quad 0 < a_1 - \kappa_1 < a_2 - \kappa_2 < \cdots < a_l - \kappa_l, \quad \kappa_1 < \kappa_2 < \cdots < \kappa_l < b.$$  

Let $\Omega_\kappa \subset M$ be the degeneracy locus defined by the conditions $\text{rk}(h : A_i \to B) \leq \kappa_i$, $i = 1, \ldots, l$, that is, the set of all points $x \in M$ where all the previous conditions are valid. Now consider the Young diagram $\{(p_1^m, \ldots, p_l^m)\}$ where

$$p_1 = a_1 - \kappa_1, \quad p_2 = a_1 - \kappa_1 - 1, \quad \ldots, \quad p_l = a_l - \kappa_l, \quad m_1 = b - \kappa_1, \quad m_2 = \kappa_1 - \kappa_1 - 1, \quad \ldots, \quad m_l = \kappa_2 - \kappa_1.$$ 

Its dual diagram is $\mu = (q_1^n, \ldots, q_l^n)$ where

$$q_1 = b - \kappa_1, \quad \ldots, \quad q_l = b - \kappa_l, \quad n_1 = a_1 - \kappa_1, \quad \ldots, \quad n_l = (a_l - \kappa_l) - (a_{l-1} - \kappa_{l-1}).$$

Set $cd(\kappa) = |\lambda|$, $s\lambda = b - \kappa_1$, $s\mu = a_l - \kappa_l$. Finally, set

$$\rho(i) = \max\{s \in [1, l] : i \leq b - \kappa_s = m_1 + \cdots + m_{l+1-s}\}, \quad i = 1, \ldots, s\lambda,$$

$$\rho'(i) = \min\{s \in [1, l] : i \leq a_s - \kappa_1 = n_1 + \cdots + n_s\}, \quad i = 1, \ldots, s\mu.$$ 

**Proposition 3** (see [Fulton 1992, 10.2]). If the codimension of $\Omega_r$ equals $cd(r)$ then the $\mathbb{Z}$-cohomology class $[\Omega_r]|_\mathbb{Z}$ of $M$ dual to $\Omega_r$ is given by

$$(8) \quad [\Omega_r]|_\mathbb{Z} = \det(c_{i, \lambda-i+j}(A_{\rho(i)}^* - B^*))_{1 \leq i, j \leq s\lambda} = \det(c_{i, \mu-i+j}(B - A_{\rho'(i)}))_{1 \leq i, j \leq s\mu},$$

where $^*$ denotes the dual bundle.

**Real case.** Consider a fixed flag of vector spaces

$$\mathbb{R}^{a_1} \subset \mathbb{R}^{a_2} \subset \cdots \subset \mathbb{R}^{a_l}.$$ 

Denote by Mat$(a_i, b)$ the space of all linear maps $\mathbb{R}^{a_i} \to \mathbb{R}^b$ identified with the space of $(a_i \times b)$-matrices. Two elements $u, v \in \text{Mat}(a_i, b)$ are called *equivalent* if for all $i = 1, \ldots, l$ the restrictions of $u$ and $v$ to $\mathbb{R}^{a_i}$ have the same rank. The set of all pairwise equivalent elements in Mat$(a_i, b)$ will be called a *stratum*. Obviously, one obtains in this way a finite stratification of Mat$(a_i, b)$. 

Using the same notation as above consider now a map \( h \) of real vector bundles

\[
h : A_1 \subset A_2 \subset \cdots \subset A_l \to B.
\]

Let \( x \in M \) be a point of the base and \( U \ni x \) be its small neighborhood such that the bundles are trivial and trivialized over \( U \). Then the map \( h \) over \( U \) is given by a family of matrices

\[
h_U : U \to \text{Mat}(a_l, b).
\]

**Definition 4.** The map \( h \) is called transversal at the point \( x \) if the map \( h_U \) is transversal to the stratum containing the point \( h_U(x) \). The map \( h \) is called transversal if it is transversal at every point \( x \in M \).

Note that the transversality condition does not depend on the trivialization of the bundles chosen over \( U \). Thom’s transversality theorem implies that a generic map \( h \) is transversal at every point \( x \).

**Corollary 5.** Let \( h \) be a transversal map. For any growth vector \( r \), its degeneracy locus \( \Omega_1 \) is a closed (possibly empty) subvariety of \( M \). The dual \( \mathbb{Z}_2 \)-cohomology class given by the intersection index with the smooth part of \( \Omega_1 \) is well-defined and given by the analog of formula (8), with the Chern classes replaced by Stiefel–Whitney classes.

**Proof.** One should follow, step by step, the proof of [Fulton 1992, Proposition 10.2]. In fact, one can show that the proof of formula (8) can be reduced to:

1. The axioms of the Chern classes, the most important of which being the Whitney formula \( c(C \oplus F) = c(E)c(F) \).
2. The isomorphism \( H^*(\mathbb{C}P^n) = \mathbb{Z}/c_1^n/1 \), where \( c_1 \) is the first Chern class of the tautological bundle over \( \mathbb{C}P^n \).
3. The construction of the Gysin map \( \phi_* : H^*(X) \to H^*(Y) \) for the proper map \( \phi : X \to Y \) of smooth manifolds, and the Gysin formula \( \phi_*(\phi^* a \cup b) = a \cup \phi_*(b) \), \( a \in H^*(X) \), \( b \in H^*(Y) \).
4. The relation \( p_*(1/(1 - c_1(S)) = c^{-1}(E^*) \),

where \( \pi : E \to M \) is any complex vector bundle, \( p : P \to M \) is its projectivization, and finally \( S \) is the natural tautological linear subbundle in the bundle \( p^\pi \) over \( P \).

All these elements have real analogs, with \( \mathbb{Z} \)-cohomology replaced by \( \mathbb{Z}_2 \)-cohomology, Chern classes by Stiefel–Whitney classes, \( \mathbb{C}P^n \) by \( \mathbb{R}P^n \) etc. \( \square \)
Application to subbundles. If we drop the restrictions (7) then for a given n-subbundle $\mathcal{V} \subset TM_m$ and a given growth vector $r = (r_1 = n \leq r_2 \leq \cdots \leq r_k = m)$ the degeneracy locus $\Sigma_r$ is the subset $\Omega_r \subset M$ for the map

$$\Phi_k : \mathcal{L}^1(\mathcal{V}) \subset \cdots \subset \bigoplus_{i=1}^{k} \mathcal{L}^i(\mathcal{V}) \rightarrow TM.$$ 

Set $L_j(\mathcal{V}) = \bigoplus_{i=1}^{j} \mathcal{L}^i(\mathcal{V})$ and $\partial(n, j) = \dim \bigoplus_{i=1}^{j} \mathcal{L}^i(\mathcal{V}) = \sum_{i=1}^{j} d(n, i)$. To apply Fulton’s formula (8) we must get rid of the redundant subspaces, i.e. those subspaces whose rank conditions are automatically satisfied due to the rank conditions imposed on the previous subspaces. (In other words, the $i$-th subspace is redundant if $\partial(n, i) - r_i = \partial(n, i - 1) - r_{i-1}$.) We define a reduced index set as a maximal subset of indices $I = (i_1, \ldots, i_l)$ for which ranks and coranks are both strictly increasing:

$$r_{i_1} < \cdots < r_{i_l} < m, \ 0 < \partial(n, i_1) - r_{i_1} < \partial(n, i_2) - r_{i_2} < \cdots < \partial(n, i_l) - r_{i_l}.$$ 

One gets the Young diagram $\lambda(r) = (p_1(r)^{m_1(r)}, \ldots, p_{l}(r)^{m_{l}(r)})$, where

$$p_1(r) = \partial(n, i_l) - r_{i_l}, \quad p_2(r) = \partial(n, i_{l-1}) - r_{i_{l-1}}, \quad \ldots, \quad p_{l}(r) = \partial(n, i_1) - r_{i_1}, \quad m_1(r) = m - r_{i_1}, \quad m_2(r) = r_{i_1} - r_{i_{l-1}}, \quad \ldots, \quad m_{l}(r) = r_{i_2} - r_{i_1}.$$ 

Its dual diagram is $\mu(r) = (q_1(r)^{n_1(r)}, \ldots, q_{l}(r)^{n_{l}(r)})$, where

$$q_1(r) = m - r_{i_1}, \quad \ldots, \quad q_{l}(r) = m - r_{i_l}, \quad n_1(r) = \partial(n, i_1) - r_{i_1}, \quad \ldots, \quad n_{l}(r) = (\partial(n, i_l) - r_{i_l}) - \partial(n, i_{l-1}) - r_{i_{l-1}}).$$

Finally, we set $cd(r) = |\lambda(r)| = |\mu(r)|$ is the area of either of these Young diagrams.

Analogously, we have $s\lambda(r) = m - r_{i_1}, \ s\mu(r) = \partial(n, i_1) - r_{i_1}$, and

$$\rho_{r}(i) = \max\{s \in [1, l] : i \leq m - r_{i_1} = m_1(r) + \cdots + m_{l+1-i}(r)\}, \quad i = 1, \ldots, s\lambda(r),$$

$$\rho'_{r}(i) = \min\{s \in [1, l] : i \leq \partial(n, i_l) - r_{i_l} = n_1(r) + \cdots + n_s(r)\}, \quad i = 1, \ldots, s\mu(r).$$

**Definition 6.** The number $cd(r)$ is called the expected codimension of $\Sigma_r$.

**Main Result.** If $\text{codim}(\Sigma_r(\mathcal{V}))$ coincides with the expected codimension $cd(r)$, then the $\mathbb{Z}_2$-cohomology class $[\Sigma_r]_{\mathbb{Z}_2}$ of the base manifold $M$ dual to $\Sigma_r$ is

$$[\Sigma_r]_{\mathbb{Z}_2} = \det(w_{\lambda(i)-i+j}(L(\mathcal{V})^{\ast}_{\rho(i)} - TM^{\ast}))_{1 \leq i, j \leq s\lambda(r)} = \det(w_{\mu(i)-i+j}(TM - L(\mathcal{V})^{\ast}_{\rho'(i)}))_{1 \leq i, j \leq s\mu(r)},$$

where the $w_j$ are the Stiefel–Whitney classes.

**Examples** (compare [Kazarian et al. 1997]). Consider a generic 2-subbundle in $TM_4$. There are three possible nonmaximal growth vectors, $(2, 2, 4)$, $(2, 3, 3)$,
and (2, 2, 3, 4). The coincidence of the actual and the expected codimensions in this case follows from the normal forms in [Zhitomirskii 1990]. (In the case (2, 2, 2, . . .) the codimension is ≥ 5.)

Case r = (2, 2, 4). The reduced index set is I = [2], i.e. we have to consider only the map \( \Phi_2 : L_2(V) \to TM \) of the usual bundles and determine the locus of points where \( r_k(\Phi_2) \leq 2 \). One has \( r_k(L_2(V)) = 3, r_k(TM) = 4, \lambda(r) = (1^2), s\lambda(r) = 2, cd(r) = 2 \). Finally, \( \mu(r) = 2, s\mu(r) = 1 \) and \( \rho'(1) = 1 \). Therefore,

\[
[S(2,2,4)]_{\underline{2}} = w_2(TM - L_2(V)) = w_2(M) + w_2(V) + w_1^2(V).
\]

Case r = (2, 3, 3). The reduced index set is I = [3], i.e. we have to consider only the map \( \Phi_3 : L_3(V) \to TM \) of the usual bundles and determine the locus of points where \( r_k(\Phi_3) \leq 3 \). One has \( r_k(L_3(V)) = 5, r_k(TM) = 4, \lambda(r) = (2^1), s\lambda(r) = 1, cd(r) = 2 \). Finally, \( \mu(r) = (1^3), s\mu(r) = 2 \) and \( \rho'(1) = 1, \rho'(2) = 1 \). Therefore,

\[
[S(2,3,3)]_{\underline{2}} = \begin{vmatrix} w_1(TM - L_3(V)) & w_2(TM - L_3(V)) \\ w_0(TM - L_3(V)) & w_1(TM - L_3(V)) \end{vmatrix} \\
= w_1^2(M) + w_1^2(V) + w_2(M) + w_1(M)w_1(V).
\]

Case r = (2, 2, 3, 4). The reduced index set is I = [2, 3], so we have to consider the map \( \Phi : L_2(V) \subset L_3(V) \to TM \). One has \( \lambda(r) = (2, 1), s\lambda(r) = 2, cd(r) = 3 \). Now, \( \mu(r) = (2, 1), s\mu(r) = 2 \) and \( \rho'(1) = 1, \rho'(2) = 2 \). Therefore,

\[
[S(2,2,3,4)]_{\underline{2}} = \begin{vmatrix} w_2(TM - L_2(V)) & w_3(TM - L_2(V)) \\ w_0(TM - L_3(V)) & w_1(TM - L_3(V)) \end{vmatrix} \\
= w_1(M)w_2(M) + w_2(M)w_1(V) + w_1^3(V) + w_3(M).
\]

These answers are obtained through standard manipulations with the total Chern class of \( TM \) and \( \Sigma^k_n \).

5. Transversality property for subbundles and general properties on \( f^*V \)

To be able to apply formula (11) to subbundles, one needs to show that a certain transversality property is valid for the map \( \Phi_k : \bigoplus_{i \leq k} \Sigma^i(V) \to TM \); see §2. This condition can be formulated as follows. The total space \( \text{Hom}(\bigoplus_{i \leq k} \Sigma^i(V), TM) \) has a natural stratification according to different degenerations of the growth vector. The transversality property says that the section of the above bundle determined by the map \( \Phi_k \) is transversal to each stratum of this natural stratification.

Naturally one wants to know if the transversality property is valid for generic \( n \)-dimensional subbundles \( \mathcal{V} \subset TM \). The conjecture stated below claims that this is indeed the case. (Up to codimension \( m - \sqrt{m} \) a similar statement is shown to be valid in [Gershkovich and Vershik 1988].)
Since the transversality property is essentially local, we formulate it in local terms.

**Local problem.** Take $M = \mathbb{R}^m$ with a fixed system of coordinates $x_1, \ldots, x_m$ and consider the set $\Omega^0$ of germs of $n$-subbundles in $\mathbb{R}^n$ such that for any $\mathcal{V} \in \Omega^0$ the subspace $\mathcal{V}(0)$ at the origin is spanned by $\partial/\partial x_1, \ldots, \partial/\partial x_n$. The set $\Omega^0$ can be identified with the set of all $n$-tuples of vector-fields $v_1, \ldots, v_n$ of the form $v_i = \partial/\partial x_i + \sum_{j=n+1}^m a_{i,j}(x_1, \ldots, x_m)\partial/\partial x_j$. Indeed, fixing the standard Euclidean structure on $\mathbb{R}^n$ we can uniquely lift the vector fields $\partial/\partial x_1, \ldots, \partial/\partial x_n$ to any subbundle $\mathcal{V} \in \Omega^0$ and get the $n$-tuple of vector fields $v_1(\mathcal{V}), \ldots, v_n(\mathcal{V})$ with the above properties.

**Remark.** For each $k$ we have the derived map $\Psi_k : \mathcal{V} \to Fl_k(\mathcal{V})$ where $Fl_k(\mathcal{V}) = (\mathcal{V} = \mathcal{V}_1 \subset \mathcal{V}_2 \subset \cdots \subset \mathcal{V}_k)$. Each $Fl_k(\mathcal{V})$ is the image under the canonical map of the $\mathcal{A}$-module $N_k(\mathcal{V})$, see §2. Fixing the standard Euclidean structure, the $n$-tuple of vector fields $\partial/\partial x_i$ and the Hall basis we obtain the standard set of sections for all $\mathcal{V} \in \Omega^0$ and in all $N_k(\mathcal{V})$. This gives us a noncanonical isomorphism between $\mathcal{V}_1$ and $\mathcal{V}_2$ and between $N_k(\mathcal{V}_1)$ and $N_k(\mathcal{V}_2)$ as $\mathcal{A}$-modules for any subbundles $\mathcal{V}_1, \mathcal{V}_2$ and any positive integer $k$. Localizing we can identify the jets of the subbundle $\mathcal{V}$ with the jets of the $n$-tuples of vector fields $v_1(\mathcal{V}), \ldots, v_n(\mathcal{V})$ and the jets of $Fl_k(\mathcal{V})$ with the jets of the $\partial(n,k)$-tuples of vector fields obtained from $v_1(\mathcal{V}), \ldots, v_n(\mathcal{V})$ by applying the commutations prescribed by the elements in the chosen Hall basis. (Recall that $\partial(n,k) = \sum_{j=1}^k d(n,i)$.) The map $\Psi_k$ induces the well-defined map $\Psi^i_k : j^{k+1}(\mathcal{V}) \to j^i(Fl_k(\mathcal{V}))$ of the corresponding jets.

**Remark.** The 0-jet of $Fl_k(\mathcal{V})$ can be represented by a $m \times \partial(n,i)$-matrix of the form

$$
\begin{pmatrix}
\vdots & n & m-n & d(n,2) & d(n,3) & \ldots & d(n,k) & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
n & 1 & 0 & 0 & 0 & \ldots & 0 & \vdots \\
m-n & 0 & 0 & * & * & \ldots & * & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots 
\end{pmatrix}
$$

(12)

Here $\mathbb{I}$ and 0 denote the identity and the zero matrices of the sizes given in the first row and the first column and $'*'$ stands for arbitrary real entries.

**Notation.** Let $\text{Mat}^0(n,m,k)$ denote the subset of all $m \times \partial(n,k)$-matrices of the above form (12) and $J(n,m,k)$ denote the space of $k$-jets of $\mathcal{V} \in \Omega^0$. The space $J(n,m,k)$ is isomorphic to the space of all $n(m-n)$-tuples of polynomials in $m$ variables of degree $\leq k$, see above. Obviously, $\dim J(n,m,k) = (m-n)n^{(m+k)}$ and $\dim \text{Mat}^0(n,m,k) = (m-n)\sum_{i=2}^k d(n,i)$. 
The main object of the remaining part of the paper is the polynomial map of affine spaces

\[ \Psi_k^0 : J(n, m, k - 1) \to \text{Mat}^0(n, m, k). \]

The space \( \text{Mat}^0(n, m, k) \) has the following natural stratification. We start with the obvious inclusions \( \text{Mat}^0(n, m, 1) \subset \text{Mat}^0(n, m, 2) \subset \cdots \subset \text{Mat}^0(n, m, k) \). Now fixing the growth vector \( r = (r_1 = n \leq r_2 \leq \cdots \leq r_k \leq \partial(n, k)) \) we define the subset \( \text{Mat}^0_k(n, m, k) \subset \text{Mat}^0(n, m, k) \) of all matrices whose restrictions to \( \text{Mat}^0_k(n, m, i) \) have rank \( \leq r_i \) for all \( i = 1, \ldots, k \). Rather obviously, the codimension of \( \text{Mat}^0_k(n, m, k) \) in \( \text{Mat}^0(n, m, k) \) equals \( cd(r) \).

Finally we are in position to formulate the required transversality property.

**Main Conjecture.** The subset of points \( x \) in \( J(n, m, k) \) such that the map \( \Psi_k^0 \) is nontransversal to the stratum containing the point \( \Psi_k^0(x) \in \text{Mat}^0(n, m, k) \) has codimension in \( J(n, m, k) \) strictly exceeding \( m \).

Thom’s transversality theorem implies that the validity of the above conjecture immediately leads to the transversality assumption for generic \( n \)-dimensional subbundles in \( TM \).

We were unable to prove the above conjecture in its complete generality but we were able to settle a number of cases given below.

**Transversality Theorem.** The required transversality property holds either for \( n \geq 3 \) and \( m \leq \frac{1}{2} n(n + 1)(2n + 1) = \partial(n, 3) \) or for \( n = 2 \) and \( m \leq 8 = \partial(2, 4) \). Namely, for \( k = 2, 3 \) and any \( m \geq n \) the map \( \Psi_k^0 \) is a submersion. Also, for \( n = 2 \) and any \( m \geq n \) and \( k = 4 \) the map \( \Psi_4^0 \) is a submersion.

**Violation of transversality for big codimensions.** We will finish this section by pointing out that the behavior of the image \( \Psi_k^0(J(n, m, k - 1)) \) w.r.t the natural rank stratification of the space \( \text{Mat}^0(n, m, k) \) is highly nontrivial. The next two statements show that one can only hope that transversality holds for the strata of relatively small codimension, as stated in the main conjecture.

**Lemma 7.** For any fixed \( m \geq n \) and for \( k > \text{const} \log_2 \left( \begin{pmatrix} 2m \\ m \end{pmatrix} \right) \), one has

\[ \dim J(n, m, k - 1) < \dim \text{Mat}^0(n, m, k), \]

and therefore \( \Psi_k \) is not surjective.

**Proof.** This is simply dimension count since

\[ \dim J(n, m, k - 1) = (m - n)n \begin{pmatrix} m + k - 1 \\ k - 1 \end{pmatrix} \]

and \( \dim \text{Mat}^0(n, m, k) = (m - n) \sum_{i=2}^{k} d(n, i) \) where \( kd(n, k) = \sum_{j\leq k} \mu(j)n^{k/j} \), where \( \mu(j) \) denotes the Möbius function. \( \square \)
Lemma 8. For any \( m \geq n \geq 3 \) and \( k \geq 4 \) the map \( \Psi_k \) is never onto. The same holds for \( n = 2 \) and \( k \geq 5 \).

Proof. Consider a matrix in \( \text{Mat}^0(n, m, 4) \) of the form

\[
\begin{pmatrix}
\vdots & n & m - n & d(n, 2) & d(n, 3) & d(n, 4) & \vdots \\
1 & 0 & 0 & * & \text{full} & \\
0 & 0 & 0 & * & \text{rank} & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{pmatrix}
\]

Such a matrix is never in the image of \( \Psi_4 \) since the fourth homogeneous component \( H^4 \) of the Hall basis contains elements of the form \( ((v_i, v_j), (v_p, v_q)) \) which vanish since the first commutators of all basic vector fields vanish. (For \( n = 2 \) the same effect happens for \( k \geq 5 \).)

6. Transversality theorem and defect vectors

This section is devoted to the proof of transversality theorem of the previous section as well as to the description of generic defect vectors in our situation. Namely, let \( w = (((v_{i_1}, v_{i_2}),(v_{i_3}, v_{i_4})) \ldots ((v_{i_{k-4}}, v_{i_{k-3}}),(v_{i_{k-1}}, v_{i_k}))) \in H^k \) be any element of length \( k \) in the Hall basis.

Definition 9. By the depth \( dp_w(v_{i_j}) \) of any variable \( v_{i_j} \) in \( w \) we denote the difference between the number of opening and closing parentheses preceding \( v_{i_j} \). By the depth \( dp(w) \) of \( w \) we mean \( \max_j dp_w(v_{i_j}) \).

Obviously, for any \( j \) one has \( dp_w(v_{i_j}) \leq dp(w) \) and there exist at least 2 different values of \( j \) for which \( dp_w(v_{i_j}) = dp(w) \).

Let \( H^p \) denote the \( p \)-th homogeneous component of \( H \) and \( H^{p,q} \) be its subset containing all elements of depth \( q \). Obviously, \( q \) can vary between \( \lfloor \log_2(p) \rfloor + 1 \) and \( p \). We restrict the map \( \Psi^0_k \) to the union \( \bigcup_{j=1}^{k} H^{j,j} \).

Lemma 10. For any \( k \) and \( m \geq n \) the restricted map

\[
\Psi^\text{res}_k : J(n, m, k - 1) \to \text{Mat}^0_{\text{res}}(n, m, k)
\]

is a submersion where \( \text{Mat}^0_{\text{res}}(n, m, k) \subset \text{Mat}^0(n, m, k) \) contains only rows corresponding to the elements from \( \bigcup_{j=1}^{k} H^{j,j} \).

Proof. We use induction on \( k \).

Base of induction: \( k = 2 \). One has \( (v_i, v_j)_l = a_{j,i,l} - a_{l,j,i} \) where \( (v_i, v_j)_l \) is the value of the \( l \)-th component of the commutator \( (v_i, v_j) \) at the origin, \( a_{j,i,l} \) is the coefficient at \( x_l \) in the \( l \)-th component of \( v_j \) and \( a_{l,j,i} \) is the coefficient at \( x_j \) in
the $l$-th component of $v_l$. Since $a_{j_1,j_2}$ and $a_{i,j_1,j_2}$ are independent parameters in the preimage the map from $J(n, m, 1)$ to $\text{Mat}^{0}_{\text{res}}(n, m, 2)$ is, obviously, a submersion.

Induction step. Assume the statement is proved for $j < k$. Elements in $H^{k,k}$ are in one-to-one correspondence with the commutators $(v_{i_1}(v_{i_2}(\ldots (v_{i_{k-1}}, v_{i_k}) \ldots)))$, where $i_1 \geq i_2 \geq \cdots \geq i_{k-1} < i_k$; see the definition of the Hall basis. In particular, the number $\#_k(n)$ of such elements equals $\sum_{j=0}^{n-1} (\binom{j+k-2}{j}) j$. For example,

\[
\begin{align*}
\#_2(n) &= \frac{1}{2} n(n-1), \\
\#_4(n) &= \frac{1}{8} (n-1) n(n+1)(n+2), \\
\#_3(n) &= \frac{1}{3} n(n-1)(n+1), \\
\#_5(n) &= \frac{1}{30} n(n-1)^2(n+2)(n+3), \\
\#_6(n) &= \frac{1}{4} n^2(n^2-1)(n^2+n+2).
\end{align*}
\]

One has

\[
(v_{i_1}(v_{i_2}(\ldots (v_{i_{k-1}}, v_{i_k}) \ldots)))_l = \text{const} + \frac{a_{i_k,i_1,i_2,\ldots i_{k-1},i}}{c(i_1,\ldots,i_{k-1})} - \frac{a_{i_{k-1},i_1,i_2,\ldots i_{k-2},i_k,i}}{c(i_1,\ldots,i_{k-2},i_k)}.
\]

Here $(v_{i_1}(v_{i_2}(\ldots (v_{i_{k-1}}, v_{i_k}) \ldots)))_l$ is the value at 0 of the $l$-th component of the commutator vector field corresponding to the element of $H^{k,k}$ under consideration. The constant summand on the right depends only on the $(k-2)$-jet of the basic vector fields $v_1(\mathcal{V}), \ldots, v_n(\mathcal{V})$, and $c(i_1,\ldots,i_{k-1}), c(i_1,\ldots,i_{k-2},i_k)$ are certain combinatorial constants. Finally, $a_{i_k,i_1,i_2,\ldots i_{k-1},i}$ denotes the variable coefficient at the product $x_{i_1}x_{i_2}x_{i_{k-1}}$ in the $l$-th component of the basic vector field $v_{i_k}$, and $a_{i_{k-1},i_1,i_2,\ldots i_{k-2},i_k,i}$ denotes the coefficient at the product $x_{i_1}x_{i_2}x_{i_{k-2}}x_{i_k}$ in the $l$-th component of $v_{i_{k-1}}$. One can easily check that $a_{i_k,i_1,i_2,\ldots i_{k-1},i}$ appears only in the commutator $(v_{i_1}(v_{i_2}(\ldots (v_{i_{k-1}}, v_{i_k}) \ldots)))_l$. Since all these variable coefficients are independent parameters in the preimage, we get a submersion of the space $\Omega_3$ on the space $\widetilde{\text{Mat}}^{0}_{\text{res}}(n, m, k)$. Here $\Omega_3$ is the subspace of all subbundles in $\Omega$ with some arbitrary fixed $(k-2)$-jet $\hat{3}$ and $\widetilde{\text{Mat}}^{0}_{\text{res}}(n, m, k)$ is the last block in $\text{Mat}^{0}_{\text{res}}(n, m, k)$.

**Corollary 11.** Lemma 10 implies the Transversality Theorem (page 247).

**Proof.** Indeed, if $n = 2, m \leq 8$ or $n \geq 3, m \leq \frac{1}{6} n(n+1)(2n+1) = \partial(n, 3)$, one has $H^k = H^{k,k}$. \hfill \square

**Description of generic defect vectors.** We conclude this section with a conjectural description of all degenerations of the associated flag bundle $f^*\mathcal{V}$ that might occur for a generic $n$-dimensional bundle $\mathcal{V}$ in $TM$.

**Definition 12.** For any growth vector $r = (r_1 = n, \ldots, r_k = m)$, the defect vector of $r$ is

\[
(\partial(n, 1) - r_1; (\partial(n, 2) - r_2) - (\partial(n, 1) - r_1); \ldots; (\partial(n, k-1) - r_{k-1}) - (\partial(n, k-2) - r_{k-2}); 0).
\]
(Up to a reduction of redundant indices this definition coincides with that of the $n_i$ in (10).) A stratum $St_r$ is called admissible if $\text{codim}St_r \leq m$, and potentially admissible if $cd(r) \leq m$. Next, $St_r$ is called bounding (potentially bounding) if it is nonadmissible (potentially nonadmissible) and is not contained in the closure of any nonadmissible (potentially nonadmissible) stratum distinct from $St_r$.

**Lemma 13.** For any $n \geq 3$ and any $k \geq 1$ one has $d(n, k + 1) > \partial(n, k) = \sum_{j=1}^{k} d(n, j)$. For $n = 2$ and any $k \geq 1$ one has $d(2, k + 1) + d(2, k + 2) > \partial(2, k) = \sum_{j=1}^{k} d(2, j)$.

**Proof.** We will consider only the case $n \geq 3$. We construct for each element $a \in H^j$, $j \leq k$ the unique element in $H^{k+1}$, i.e. embed $\bigcup H_{j=1}^{j=1} \cup \ldots \cup H_{j=k}^{j=k}$ into $H^{k+1}$, thus proving that $d(n, k + 1) > \partial(n, k)$. Recall that the Hall family is linearly ordered and each element in the Hall family has a unique representation in the form $(a(bc))$ where $a, b, c$ satisfy the conditions: $a, b, c, bc$ belong to $H$, $a \geq b$ and $b < c$. For all $j < \frac{k+1}{2}$ we associate to any element $a \in H^j$ the element $(a(v_1(v_1(\ldots (v_1, v_2)\ldots )))a) \in H^{k+1}$. Now let $j > \frac{k+1}{2}$ and $a$ be some element in $H^j$. Then $a = (f, g)$ where $f < g$. Assume additionally that $2 \text{lng}(f) + \text{lng}(g) \leq k + 1$. Then we associate to $a$ the element $(h(f, g))$ where $h$ is the maximal element in $H^{k+1-\text{lng}(h) - \text{lng}(g)}$. (Note that, by definition, $h \geq f$.) If $2 \text{lng}(f) + \text{lng}(g) > k + 1$ then we choose for each $a = (f, g) \in H^j$ the element $(f(h, g))$ where $h$ is the minimal element in $H^{k+1-\text{lng}(h) - \text{lng}(g)}$. The last case to consider is $H^j$ when $k = 2l - 1$. One has that $\Lambda^2(H^j)$ is embedded into $H^{j+1}$ and under the assumption $n \geq 3$ one has $\text{dim} \Lambda^2(H^j) \geq \text{dim} H^j$. Combining all choices together we obtain a set-theoretic embedding of $\bigcup_{j=1}^{k} H^j$ into $H^{k+1}$. The result follows. More detailed consideration shows that $d(2, k + 1) + d(2, k + 2) > \partial(2, k)$. □

**Remark.** The transversality property is equivalent to showing that the codimension of each potentially admissible (resp. potentially bounding) stratum $St_r$ equals $cd(r)$ and therefore $St_r$ is, in fact, admissible (resp. bounding).

We now list the defect vectors of all potentially admissible strata.

**Lemma 14.** For $n \geq 3$ the defect vectors of all potentially admissible strata are as in one of the following cases. We assume that $\partial(n, p) \leq m < \partial(n, p + 1)$.

(a) There is a $1$ in position $l < p$, a $\chi$ in position $p$, and $0$ in all other positions, with the further restrictions $(m - \partial(n, p) + 1 + \chi) \chi \leq \partial_l - 1$ and $\chi \geq 0$.

(b) There is a $1$ in position $l < p$, a $\chi$ in position $p + 1$, and $0$ in all other positions, with the further restrictions $(m - \partial(n, p + 1) + 1 + \chi) \chi \leq \partial_l - 1$ and $\chi + 1 + m - \partial(n, p + 1) \geq 0$.

(c) There is a $\chi$ in position $p$, a $\nu$ in position $p + 1$, and $0$ in all other positions, with the further restrictions $(m - \partial(n, p) + \chi) \chi + (m - \partial(n, p + 1) + \chi + \nu) \nu \leq m$ and $m - \partial(n, p + 1) + \chi + \nu \geq 0$. 
Proof: If the rank of the image of some $L_l(V) = \bigoplus_{i=1}^l \mathcal{L}^i(V)$, where $l < p$, drops then it can drop exactly by 1. Indeed, assume that it drops by at least 2 then using (10) we get for the dual diagram $\mu_r$ that $q_1 \geq m - \partial(n, l) + 2$ and $n_1 \geq 2$. But, by Lemma 13 one gets $2(m - \partial(n, l) + 2) > m$ which contradicts the assumption $cd(r) \leq m$. Analogously, if $L_l(V)$, with $l < p$, has corank 1, the corank can possibly increase again only for either $L_p(V)$ or $L_{p+1}(V)$; see description below. Indeed, assume that the corank drops by 1 at $L_l(V)$ and by 2 at $L_{l_1}(V)$ where $l_1 < l < k$. Then by (10) one has $q_1 = m - \partial(n, l_1) + 1, n_1 = 1, q_2 = m - \partial(n, l_2) + 2, n_2 = 1$. But, again by Lemma 13, one gets $q_1 + q_2 > m$, which contradicts $cd(r) \leq m$. Therefore the second rank drop can only occur either at position $p$ or $p + 1$. If we have that $L_l(V), l < p$ has corank 1 then further corank drops at both positions $p$ and $p + 1$ are simultaneously impossible, i.e., only one extra drop is allowed.

Indeed, assume that we have coranks 1, 2, 3 at positions $l$, $p$, $p + 1$ respectively. Then by (10) one gets $q_1 = m - \partial(n, l) + 1, n_1 = 1, q_2 = m - \partial(n, p) + 2, n_2 = 1$; $q_3 = 1, n_3 = \partial(n, p + 1) - m - 1$. But $q_2n_2 + q_3n_3 = m - \partial(n, p) + 2 + \partial(n, p + 1) - m - 1 = d(n, p) + 1$. Again, by Lemma 13 one gets $q_1n_1 + q_2n_2 + q_3n_3 > m$. Thus if the rank drops in some position prior to $p$ we are left with cases (a) and (b). The above list of inequalities follows from the expressions for the terms in the dual Young diagram $\mu = (q_1^{n_1}, q_2^{n_2})$ given below. In case (a) one gets $q_1 = m - \partial(n, l) + 1, n_1 = 1, q_2 = m - \partial(n, p) - 1 - \chi, n_2 = \chi$. In case (b) one gets $q_1 = m - \partial(n, l) + 1, n_1 = 1, q_2 = m - (\partial(n, p + 1) - 1 - \chi), n_2 = \chi$. Finally, in case (c) one gets $q_1 = m - \partial(n, p) + \chi, n_1 = \chi, q_2 = m - \partial(n, p + 1) + \chi + \nu, n_2 = \nu$.

The inequalities express the condition $cd(r) \leq m$ and the condition that the second rank drop actually occurs. One can easily show that (a) are (b) are mutually exclusive: for a given pair $(n, m)$ either the inequalities for (a) or those for (b) can be satisfied under the assumption that $\chi > 0$.

**Corollary 15.** The defect vectors for all potentially bounding strata have one of the forms given in the table.

<table>
<thead>
<tr>
<th>$l_1 - 1$</th>
<th>$l_1$</th>
<th>$l_1 + 1$</th>
<th>$l_2 - 1$</th>
<th>$l_2$</th>
<th>$l_2 + 1$</th>
<th>$p - 1$</th>
<th>$p$</th>
<th>$p + 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>(b)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>(c)</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>$\chi$</td>
</tr>
<tr>
<td>(d)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>(e)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>...</td>
<td>...</td>
<td>0</td>
</tr>
</tbody>
</table>
In cases (a)–(c), whose form coincides with that in Lemma 14, we additionally require that the inequalities from the formulation of Lemma 14 are violated for the considered values of $\nu$ and $\chi$ but satisfied for any smaller nonnegative values of these variables.

Proof. Obtained by a simple case-by-case consideration. \qed

7. Final remarks

1. The basic problem related to the above topological obstructions is to what extent vanishing of these obstructions guarantees the existence of a maximal growth subbundle. The result of T. Vogel on the existence of Engel structures on parallelizable 4-manifolds brings a certain amount of optimism about this problem. For example:

Problem. Does every closed parallelizable $m$-dimensional manifold admit a maximal growth distribution of rank $1 < n < m$?

2. The theory developed here is incomplete since we lack a proof of the transversality conjecture. Essentially, to accomplish the proof one should only consider the defect vectors listed at the end of Section 6. The authors are convinced that progress in this direction will be intimately related to a detailed study of combinatorics of different Hall bases.

3. The natural representation of $\text{GL}_n(\mathbb{C})$ in $\mathcal{L}_n^l$ was extensively studied in the series of papers [Thrall 1942; Wever 1949; Klyachko 1974] and some others. There exists an interesting formula for the multiplicity of each irreducible representation (corresponding to some Young diagram $\mu$) in $\mathcal{L}_n^l$ analogous to the one for the dimension of $\mathcal{L}_n^l$. This suggests the existence of a much more sophisticated theory of characteristic classes for subbundles in the tangent bundle since there exist many different natural filtrations in $\mathcal{L}_n$ the most complicated of which coming from the direct sum of irreducible $\text{GL}_n(\mathbb{C})$-representations. The first step in this direction will be to find analogs of Theorem 2 for these other filtrations; the second, to find the analogous transversality theorems. As an example of characteristic classes different from the ones considered above one can try to calculate the characteristic classes related to the depth filtration of $\mathcal{L}_n$ introduced in Section 6. For this case there exists a character formula in terms of plethysms analogous to (5) suggested to the authors by C. Reutenauer but it is unfortunately rather complicated.

4. Another natural question related to the transversality theorem is to understand for strata of what corank in the space $\text{Mat}^0(n, m, k)$ it holds and therefore to extend the determinantal formulas to the case of nongeneric subbundles or families of subbundles.

Problem. Generalize the transversality theory to the case of nongeneric subbundles.
8. Appendix

Here is a Mathematica program for the calculation of Chern classes of homogeneous components $L^k$ of free Lie algebra bundles up to a given order.

order=4;

(* This function computes the Chern character via the total Chern class *)
classtochar[cc_]:= (resc[n-t D[Log[cc],t],-1]/.t^k_->t^k/k!)
+O[t]^(order+1)//ExpandAll;

(* This function calculates total Chern class via Chern character *)
chartoclass[ch_] := Exp[-Integrate[PolynomialQuotient[resc[ch,-1]/.
t^k_->t^k k!,t],t]+O[t]^(order+1)]//ExpandAll;

(* This function makes the rescaling $\eta \rightarrow (\eta)_l$. *)
resc[eta_,l_]:=Normal[eta]/.t->l t;

(* Total Chern class of original bundle. *)
class=1+Sum[c[i]t^i,i,order]+O[t]^(order+1);

(* Chern character of original bundle. *)
char=classtochar[class];

(* Computes the Chern character of $L^k$ using Theorem 2. *)
charoFL[k_]:= (Plus@((MoebiusMu[#]resc[char^(-k/#),#]/k)/@
(Divisors[k])))//ExpandAll;

(* Total Chern class of $L^k$ as a series in $t$ *)
classoFL[k_]:=chartoclass[charoFL[k]];

Acknowledgments

The starting point of this article was the special case of 2-subbundles on $M_4$ considered in [Kazarian et al. 1997]. Later the authors realized that analogous computations over $\mathbb{Z}_2$ can be carried out in the general setup described in the present paper. Sincere thanks are due to R. Montgomery for important discussions of the subject and numerous remarks which substantially improved the quality of exposition. Kazarian is grateful to IHES (Paris), MPIM (Bonn) and MSRI (Berkeley), where parts of this project were carried out.

References


Received May 13, 2005. Revised August 10, 2005.

**Boris Shapiro**
**Department of Mathematics**
**Stockholm University**
**SE-106 91, Stockholm**
**Sweden**

shapiro@math.su.se

**Maxim Kazarian**
**Steklov Mathematical Institute**
**42 Vavilova St.**
**117966, Moscow GSP-1**
**Russia**

kazarian@mccme.ru
Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use \LaTeX, but papers in other varieties of \TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as \LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of Bib\LaTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal’s preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.