Pacific Journal of Mathematics

Volume 211 No. 1 September 2003

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

Founded in 1951 by

E. F. Beckenbach (1906–1982)   F. Wolf (1904–1989)

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

Regular subscription rate for 2006: $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-PROFIT CORPORATION
Typeset in LaTeX

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

BESSEL FUNCTIONS FOR GL(3) OVER A p-ADIC FIELD

Ehud Moshe Baruch

We attach Bessel functions to generic representations of GL$(3,k)$ where k is a p-adic field and study their asymptotics.

1. Introduction and main results.

Let k be a non-archimedean local field. Bessel functions for the double cover of $\mathrm{GL}(2,k)$ were constructed by Gelbart and Piatetski-Shapiro in [**G-PS**] and used to compute some epsilon factors. A more detailed analysis of the Bessel functions for $\mathrm{GL}(2,k)$ was obtained by Soudry in [**S**] and again used to compute some epsilon factors. Bessel functions for $\mathrm{GL}(2,\mathbf{R})$ were defined and analyzed by Cogdell and Piatetski-Shapiro in [**C-PS1**]. Bessel functions were also discussed in Averbuch's thesis [**Av**]. In [**Ba**] we attached Bessel functions to every generic representation of a quasi-split reductive group over k using a distribution approach similar to Harish-Chandra's approach for the character functions. In the present paper we attach Bessel functions to generic representations of $\mathrm{GL}(3,k)$ using a different method and study their asymptotics. In [**Ba2**] we use the results of this paper to show that both approaches yield the same Bessel functions.

1.1. Main results. We state here our main theorems. We shall only consider here the main Bessel function of a representation which is the one attached to the open Bruhat cell. Other Bessel functions are described in Section 6.

Let $G = \mathrm{GL}(3,k)$ and let B be the Borel subgroup of upper triangular matrices, A the subgroup of diagonal matrices and N the subgroup of upper unipotent matrices. Let ψ be a nondegenerate character of N. Let $\mathbb{W} = N(A)/A$ be the Weyl group where $N(A)$ is the normalizer of A. We identify $\mathbb{W}$ with the set of six permutation matrices in $N(A)$. Let

$$w_0 = \begin{pmatrix} & & 1 \\ & 1 & \\ 1 & & \end{pmatrix}$$

be the longest Weyl element in $\mathbb{W}$. Let (π, V) be an irreducible admissible representation of G. We say that π is generic if there exist a nonzero functional $L : V \to C$ such that

$$L(\pi(n)v) = \psi(n)L(v) \quad n \in N,\ v \in V.$$

It is well-known that such a functional is unique up to scalar multiples. We call this functional a ψ Whittaker functional. Now define

$$W_v(g) = L(\pi(g)v) \qquad v \in V,\ g \in G, \tag{1.1}$$

and let G act on the space of these functions by right translations. Then the map $v \to W_v$ gives a realization of π on a space of Whittaker functions satisfying

$$W_v(ng) = \psi(n)W_v(g) \quad n \in N,\ g \in G.$$

We denote this space by $\mathcal{W}(\pi,\psi)$. In Section 4 we define the subspace $\mathcal{W}^0(\pi,\psi)$ of $\mathcal{W}(\pi,\psi)$. In the case where π is supercuspidal we have that $\mathcal{W}^0(\pi,\psi) = \mathcal{W}(\pi,\psi)$. Let α_1, α_2 be the positive roots realized as functions on A (See (2.9)). Let $M > 0$ be a constant and let

$$A^M = A^M(w_0) = \{a \in A :\ \alpha_i(a) < M, i = 1,2\}. \tag{1.2}$$

Our first main theorem is the following:

Theorem 1.1. *Let $W \in \mathcal{W}^0(\pi,\psi)$ and M a positive constant. Then the function*

$$(a,n) \mapsto W(aw_0n)$$

defined on the set $A^M \times N$ is compactly supported in N. That is, if $W(aw_0n) \neq 0$ and $a \in A^M, n \in N$ then n is in some compact set independent of a.

Since A^M cover A as $M \to \infty$ we get the following Corollary:

Corollary 1.2. *Let π be a supercuspidal representation of G and $W \in \mathcal{W}(\pi,\psi)$ a Whittaker function associated to π. Fix $g \in Bw_0B$. Then the function*

$$n \mapsto W(gn)$$

is compactly supported in N.

Proof. Write $g = n_1aw_0n_2$ and choose M large enough such that $a \in A^M$. Since $W(gn) = \psi(n_1)W(aw_0n_2n)$ the result follows from Theorem 1.1. $\square$

This result allows us to define Bessel functions for supercuspidal representations. In order to treat all irreducible admissible representations we will need the following result which allows us to move from $\mathcal{W}(\pi,\psi)$ to $\mathcal{W}^0(\pi,\psi)$.

Theorem 1.3. *Let $W \in \mathcal{W}(\pi,\psi)$. There exist a compact open subgroup $N_0 = N_0(W)$ of N such that the function $W_{N_0,\psi} \in \mathcal{W}^0(\pi,\psi)$.*

Here $W_{N_0,\psi}$ is defined by

$$W_{N_0,\psi}(g) = \int_{N_0} W(gn)\psi^{-1}(n)dn.$$

Corollary 1.4. $\mathcal{W}^0(\pi,\psi) \neq \{0\}$.

Proof. Let $W \in \mathcal{W}(\pi,\psi)$ be such that $W(e) \neq 0$. Then $W_{N_0,\psi}(e) \neq 0$ for every compact open subgroup N_0 in N. $\square$

Let $N_1 \subset N_2 \subset N_3 \subset \dots$ be a filtration of N with compact open subgroups N_i, $i = 1, 2, \dots$, such that $N = \bigcup_{i=1}^{\infty} N_i$. We denote this filtration by $\mathcal{N}$. Let $f : N \to \mathbb{C}$ be a locally constant function.

Definition 1.5.

$$\int_N^{\mathcal{N}} f(n)dn = \lim_{m\to\infty} \int_{N_m} f(n)dn$$

if this limit exists.

Corollary 1.6. *Let $\mathcal{N} = \{N_i, i > 0\}$ be a filtration of N as above. Let $g \in Bw_0B$ and $W \in \mathcal{W}(\pi, \psi)$. Then*

$$\int_N^{\mathcal{N}} W(gn)\psi^{-1}(n)dn$$

is convergent, and the value is independent of the choice of filtration $\mathcal{N}$.

Proof. Let $N_0 = N_0(W)$ be a compact open subgroup of N as in Theorem 1.3. There exist an integer M such that $N_0 \subset N_m$ for all $m > M$. Let $m > M$. We have

$$\frac{1}{\mathrm{vol}(N_0)} \int_{N_m} W_{N_0,\psi}(gn)\psi^{-1}(n)dn = \frac{1}{\mathrm{vol}(N_0)} \int_{N_m} \int_{N_0} W(gn_1n_2)\psi^{-1}(n_1n_2)dn_1dn_2. \tag{1.3}$$

Applying Fubini and changing variables $n = n_1n_2$ we get that the last integral is the same as

$$\int_{N_m} W(gn)\psi^{-1}(n)dn.$$

By Theorem 1.1 the function $n \mapsto W_{N_0,\psi}(gn)$ is compactly supported in N, hence we can take the limit $m \to \infty$ in (1.3) to get the value

$$\frac{1}{\mathrm{vol}(N_0)} \int_N W_{N_0,\psi}(gn)\psi^{-1}(n)dn.$$

It is clear that this value is independent of the filtration $\mathcal{N}$. □

Let $g \in Bw_0B$ and define the linear functional $L_g : V \to \mathbb{C}$ by

$$L_g(v) = \int_N^{\mathcal{N}} W_v(gn)\psi^{-1}(n)dn.$$

It is easy to see that L_g is a Whittaker functional, hence it follows from the uniqueness of Whittaker functionals that there exist a scalar $j_{\pi,\psi}(g)$ such that

$$L_g(v) = j_{\pi,\psi}(g)L(v) \tag{1.4}$$

for all $v \in V$. We call $j_\pi = j_{\pi,\psi}$ the Bessel function of π. (See Section 6 for other Bessel functions.) The Bessel function j_π is defined on the set Bw_0B and we will prove that it is locally constant there. It is clear that $j_\pi(g)$ satisfies

$$j_\pi(n_1gn_2) = \psi(n_1n_2)j_\pi(g), \quad n_1, n_2 \in N,\, g \in Bw_0B,$$

hence it is determined by its values on the set Aw_0. As is the case with the character of the representation [**HC**], the Bessel function j_π is expected to be locally integrable on G. Harish-Chandra's proof of the local integrability of the character depended on certain relations between the asymptotics of the character and certain orbital integrals. In this paper we establish that the asymptotics of j_π are the same as the asymptotics of certain orbital integrals which were studied by Jacquet and Ye [**J-Y1**]. This result will allow us to show in [**Ba2**] that j_π is locally integrable and gives the Bessel distribution on G. We now describe the relation between the Bessel functions and orbital integrals.

Let $C_c^\infty(G)$ be the space of locally constant and compactly supported functions on G. Let Z be the center of G and let ω be a quasi character on Z.

For $\phi \in C_c^\infty(G)$ and $g \in Bw_0B$ we define the orbital integral (see [**J-Y1**] (6))

$$J_{\psi,\omega}(g,\phi) = \int_{N\times Z\times N} \phi(n_1zgn_2)\psi^{-1}(n_1n_2)\omega^{-1}(z)dn_1dn_2dz.$$

It follows from [**J-Y1**] that this integral converges absolutely and defines a locally constant function on Bw_0B.

Theorem 1.7. *Let π be an irreducible admissible representation of G with central character ω_π. Let $x \in G$. There exist a neighborhood U_x of x in G and a function $\phi \in C_c^\infty(G)$ such that*

$$j_{\pi,\psi}(g) = J_{\psi,\omega_\pi}(g,\phi)$$

for all $g \in U_x$.

Remark 1.8. Since $j_{\pi,\psi}$ and J_{ψ,ω_π} are only defined on Bw_0B we are really asserting the equality on $Bw_0B \cap U_x$. Another option is to define these functions as having value zero outside of Bw_0B in which case the equality above does hold. In any case, the equality is true up to a set of measure zero.

Corollary 1.9. *If $g \mapsto J_{\psi,\omega_\pi}(g,\phi)$ is locally integrable as a function on G for every $\phi \in C_c^\infty(G)$ then $j_{\pi,\psi}$ is locally integrable on G.*

Hence the question of local integrability of the Bessel function reduces to the question of the local integrability of the orbital integral. As mentioned above, we shall consider this question in [**Ba2**].

Our paper is divided as follows: In Section 2 we introduce some notations and study certain root and weight spaces. In Section 3 we describe the method of proof used for our main results and prove a result which is needed later using this method. In Section 4 we prove a more general version of Theorem 1.1. In Section 5 we prove Theorem 1.3. In Section 6 we define Bessel functions including Bessel functions attached to other Weyl elements. In Section 7 we prove Theorem 1.7.

2. Notations and preliminaries.

Let k be a non-archimedean local field. Let O be the ring of integers in k and let P be the maximal ideal in k. Let ϖ be a generator of P. We denote by $|x|$ the normalized absolute value of $x \in k$. Let $q = |O/P|$ be the order of the residue field of k. Then $|\varpi| = q^{-1}$. Let $G = \mathrm{GL}(3,k)$ and let A be the group of diagonal matrices in G consisting of matrices of the form

$$d(a_1, a_2, a_3) = \begin{pmatrix} a_1 & & \\ & a_2 & \\ & & a_3 \end{pmatrix}.$$

We let

$$Z = Z(G) = \{d(a,a,a) : a \in k^*\}. \tag{2.1}$$

Let $X(A) = \mathrm{Hom}_k(A,k)$ be the group of rational homomorphisms. Then each $\alpha \in X(A)$ is of the form

$$\alpha(d(a_1,a_2,a_3)) = a_1^{n_1} a_2^{n_2} a_3^{n_3}$$

with $n_1, n_2, n_3 \in \mathbf{Z}$. We view $X(A)$ as a group under addition where the addition is defined by

$$(\alpha+\beta)(a) = \alpha(a)\beta(a), \quad \alpha, \beta \in X(A), a \in A. \tag{2.2}$$

We let $|X| = X(A) \otimes \mathbf{R}$. Then we shall identify $|X|$ with the vector space of functions $|\alpha| = |\alpha|_{\lambda_1,\lambda_2,\lambda_3}$ from A to $\mathbf{R}$ of the form

$$|\alpha|(d(a_1,a_2,a_3)) = |a_1|^{\lambda_1} |a_2|^{\lambda_2} |a_3|^{\lambda_3} \tag{2.3}$$

where $\lambda_1, \lambda_2, \lambda_3 \in \mathbf{R}$. Here addition is defined as in (2.2) and scalar multiplication is defined by

$$(\lambda|\alpha|)(a) = (|\alpha|(a))^{\lambda}, \quad |\alpha| \in |X|, a \in A, \lambda \in \mathbf{R}.$$

We define an inner product on $|X|$ by

$$\langle \alpha_{\lambda_1,\lambda_2,\lambda_3}, \alpha_{\gamma_1,\gamma_2,\gamma_3} \rangle = \sum_{i=1}^{3} \lambda_i \gamma_i. \tag{2.4}$$

For $i, j \in \{1, 2, 3\}$, $i \neq j$ we let $\alpha_{i,j} : A \to k$ be the functions defined by

$$\alpha_{i,j}(d(a_1, a_2, a_3)) = \frac{a_i}{a_j}$$

and $|\alpha|_{i,j}(a) = |\alpha_{i,j}(a)|$. Let $\Phi = \{\alpha_{i,j}\}$. Then $|\Phi| = \{|\alpha|_{i,j}\}$ is a root system in $|X|$. We have that $\Phi = \Phi^+ \cup \Phi^-$ where $\Phi^+ = \{\alpha_{i,j} : i < j\}$ is the set of positive roots and Φ^- is the set of negative roots. Let $E_{i,j}$ be the matrix whose (i, j)th entry is 1 and all other entries are zero. For $\alpha = \alpha_{i,j} \in \Phi$ and for $b \in k$ we let

$$x_\alpha(b) = x_{i,j}(b) = I + bE_{i,j}$$
$$h_\alpha(b) = h_{i,j}(b) = bE_{i,i} - b^{-1}E_{j,j}.$$

For each $\alpha \in \Phi$ we let $N_\alpha = \{x_\alpha(b) : b \in k\}$. Let $\mathbb{W} = N(A)/A$ be the Weyl group of G. We shall identify $\mathbb{W}$ with S_3, the symmetric group. In particular if $\sigma \in S_3$ then we let w_σ be the associated permutation matrix. For every $\alpha = \alpha_{i,j} \in \Phi$ we set $w_\alpha = w_{(i,j)}$ where (i, j) is a transposition in S_3. $\mathbb{W}$ acts on Φ and $|\Phi|$ in a natural way. We have that if $i \neq j$ then

$$ax_{i,j}(b)a^{-1} = x_{i,j}(\alpha_{i,j}(a)b), \quad a \in A, b \in k \tag{2.5}$$

$$x_{i,j}(b)x_{j,i}(-b^{-1})x_{i,j}(b) = w_{(i,j)}h_{j,i}(b), \quad b \in k. \tag{2.6}$$

Let N be the subgroup of upper unipotent matrices. Then every $n \in N$ can be written uniquely in the form

$$n = x_{1,3}(b_1)x_{1,2}(b_2)x_{1,3}(b_3)$$

where $b_1, b_2, b_3 \in k$.

2.1. Roots, weights and Weyl elements. The root system $|\Phi|$ spans a subspace $|V|$ of $|X|$ given by

$$|V| = \{|\alpha|_{\lambda_1,\lambda_2,\lambda_3} : \lambda_1 + \lambda_2 + \lambda_3 = 0\}. \tag{2.7}$$

Then $\Delta = \{|\alpha|_{1,2}, |\alpha|_{2,3}\}$ is a basis for $|V|$ consisting of simple roots. If $\mathcal{B}$ is a basis for $|V|$ then we denote by $\mathcal{B}^*$ the dual basis (up to scalars) with respect to (2.4). In particular, the fundamental weights λ_1, λ_2 give the basis dual to Δ where

$$\lambda_1 = |\alpha|_{2,-1,-1}, \quad \lambda_2 = |\alpha|_{1,1,-2}. \tag{2.8}$$

We write $\Delta = \{\alpha_1, \alpha_2\}$ and $\Delta^* = \{\lambda_1, \lambda_2\}$ with

$$\alpha_1 = |\alpha|_{1,2} = |\alpha|_{1,-1,0}, \quad \alpha_2 = |\alpha|_{2,3} = |\alpha|_{0,1,-1} \tag{2.9}$$

and λ_1, λ_2 as in (2.8). If $S \subseteq \Delta$ then we let

$$\mathcal{B}(S) = \{\gamma_1, \gamma_2\}$$

where γ_i is defined by

$$\gamma_i = \begin{cases} \alpha_i & \text{if } \alpha_i \in S; \\ \lambda_i & \text{if } \alpha_i \notin S. \end{cases} \tag{2.10}$$

We let $\mathcal{B}^*(S) = (\mathcal{B}(S))^*$ (up to scalars). We shall fix the following sets:
(2.11)

$$\mathcal{B}(\emptyset) = \Delta^*, \quad \mathcal{B}(\{\alpha_1\}) = \{\alpha_1, \lambda_2\}, \quad \mathcal{B}(\{\alpha_2\}) = \{\lambda_1, \alpha_2\}, \quad \mathcal{B}(\Delta) = \Delta$$
$$\mathcal{B}^*(\emptyset) = \Delta, \quad \mathcal{B}^*(\{\alpha_1\}) = \{\alpha_1, \lambda_2\}, \mathcal{B}^*(\{\alpha_2\}) = \{\lambda_1, \alpha_2\}, \quad \mathcal{B}^*(\Delta) = \Delta^*.$$

For each $\alpha \in \Phi$ we let $w_\alpha \in \mathbb{W}$ be the reflection associated with α. In particular, $w_{\alpha_{1,2}} = w_{(1,2)}$ and $w_{\alpha_{2,3}} = w_{(2,3)}$. Since $\mathbb{W}$ is generated by $w_{(1,2)}$ and $w_{(2,3)}$ it follows that each w can be written as a product of these transpositions. We define $S(w) \subset \Delta$ to be

$$S(w) = \{\beta_1, \ldots, \beta_k\} \tag{2.12}$$

where $w = w_{\beta_1} \cdots w_{\beta_k}$ is a minimal decomposition of w into a product of such transpositions. In particular we have

$$S(e) = \emptyset, \quad S((1,2)) = \{\alpha_{1,2}\}, \quad S((2,3)) = \{\alpha_{2,3}\}$$
$$S((1,2,3)) = \Delta, \quad S((1,3,2)) = \Delta, \quad S((1,3)) = \Delta.$$

It is well-known that $S(w)$ is independent of the minimal decomposition. We define

$$S^-(w) = \{\alpha \in \Phi^+ \,:\, w(\alpha) < 0\}, \quad S^+(w) = \{\alpha \in \Phi^+ \,:\, w(\alpha) > 0\} \tag{2.13}$$

and

$$N_w^- = \prod_{\alpha \in S^-(w)} N_\alpha, \quad N_w^+ = \prod_{\alpha \in S^+(w)} N_\alpha. \tag{2.14}$$

Then $N = N_w^+ N_w^-$.

3. Method of proof.

The main method of proof in this paper is a careful analysis of the relation between the Bruhat Decomposition and the Iwasawa decomposition of our group. Such relations were explored by previous authors such as [**Bo-HC**], [**Bo**], [**J-PS-S**]. In this paper we present an explicit method of obtaining such information which follows a simple pattern. The idea is to analyze the Bruhat cells inductively going from the closed cell up to the open cell. The induction is on the length of the respective Weyl element. Another induction takes place inside an individual cell where we "peel" the root groups one by one. We find it very striking that the three main results in this paper were all proved using this method. In this section we present a new proof for a well-known result which demonstrates best how the method works.

Since we have only four different lengths for the Weyl elements of GL(3) we shall not employ induction in this paper. The induction process will be used in a future paper on $\mathrm{GL}(n)$. Before we go to the main result of this section we shall indicate how the method works.

3.1. Example. Let

$$w = w_{(1,2,3)} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then every $g \in BwB$ can be written uniquely in the form

$$g = nawx_{1,3}(b_1)x_{2,3}(b_2),$$

with $n \in N, a \in A$ and $b_1, b_2 \in k$. Our "peeling" proceeds in the following way: If we have a function F which is smooth on the right then for large enough $|b_2|$ we have that $F(gx_{3,2}(-b_2^{-1})) = F(g)$ for all $g \in G$. In particular, if $g \in BwB$ is written as above then $g_1 = gx_{3,2}(-b_2^{-1}) \in Bw_1B$ with Bw_1B being a smaller cell than BwB, that is, the length of w_1 is less than the length of w. At this point we know that the values of F on such g for which $|b_2|$ is large are determined by the values of F on smaller cells and we can use the induction hypothesis (whatever that may be for each proof) to conclude the result for such b_2. To continue, we need to control the values of F for bounded $|b_2|$ which depends on the result we are trying to prove. Once this is complete we have "peeled" the root subgroup $\{x_{2,3}(b_2)\}$ and we are left to treat the case $g = nawx_{1,3}(b_1)$. At this point we look at the case where $|b_1|$ is large and multiply g by $x_{1,3}(-b_1^{-1})$ to get $g_1 = gx_{1,3}(-b_1^{-1}) \in Bw_2B$ with length of w_2 less than the length of w. The argument continues in this way.

3.2. Iwasawa decomposition. Let $K = \mathrm{GL}(2, O)$. Then it is well-known that

$$G = NAK.$$

For every $|\alpha| \in |X|$ we extend $|\alpha|$ (see [**J-PS-S**]) to G by defining

$$|\alpha|(g) = |\alpha|(a) \tag{3.1}$$

where $g = nak$, $n \in N, a \in A, k \in K$, is an Iwasawa decomposition of g. It is easy to see that $|\alpha|$ is independent of the choice of decomposition.

Recall that $\Delta^* = \{\lambda_1, \lambda_2\}$ is the set of fundamental weights where $\lambda_1 = |\alpha|_{2,-1,-1}$ and $\lambda_2 = |\alpha|_{1,1,-2}$ (See (2.3).) We view λ_1 and λ_2 as functions on G as above. The main theorem of this section is the following:

Theorem 3.1. *Let $\lambda \in \Delta^*$ and $w \in \mathbb{W}$. Then*

$$\lambda(wn) \le 1$$

for every $n \in N_w^-$.

Proof. The proof is a case by case argument starting with the smaller Bruhat cells and moving to the larger cells.

Case 1. $w = e$.
In this case $N_w^- = \{e\}$. Hence, $wn = e$. Since $\lambda(e) = 1$ we are done.

Case 2. $w = w_{(1,2)}$.
In this case $N_w^- = N_{1,2}$. Hence $wn = wx_{1,2}(b)$. If $|b| \le 1$ then $wn \in K$ and $\lambda(wn) = 1$. If $|b| > 1$ then $x_{2,1}(-b^{-1}) \in K$ and

$$\lambda(wn) = \lambda(wx_{1,2}(b)) = \lambda(wx_{1,2}(b)x_{2,1}(-b^{-1})) = \lambda(h_{2,1}(b)) \le 1.$$

Case 3. $w = w_{(2,3)}$.
This case is similar to Case 2 and is omitted.

Case 4. $w = w_{(1,2,3)}$.
In this case $N_w^- = N_{1,3}N_{2,3}$. We first look at a special case where $wn = wx_{1,3}(b_1)$. If $|b_1| \le 1$ then $wn \in K$ and $\lambda(wn) = 1$. If $|b_1| > 1$ then $x_{3,1}(-b_1^{-1}) \in K$ and

$$\begin{aligned}\lambda(wx_{1,3}(b_1)) &= \lambda(wx_{1,3}(b_1)x_{3,1}(-b_1^{-1})) \\ &= \lambda(x_{2,3}(1)h_{2,1}(b_1)w_{(2,3)}) = \lambda(h_{2,1}(b_1)) \le 1.\end{aligned}$$

For the general case, assume

$$wn = wx_{1,3}(b_1)x_{2,3}(b_2).$$

If $|b_2| \le 1$ then $x_{2,3}(b_2) \in K$ and $\lambda(wn) = \lambda(wx_{1,3}(b_1)) \le 1$ by the special case above. If $|b_2| > 1$ then $x_{3,2}(-b_2^{-1}) \in K$ and we have

$$\lambda(wn) = \lambda(wx_{1,3}(b_1)x_{2,3}(b_2)x_{3,2}(-b_2^{-1})).$$

Since $wx_{1,3}(b_1)x_{2,3}(b_2)x_{3,2}(-b_2^{-1}) = n'h_{3,1}(b_2)w_{(1,2)}n''$ for some $n' \in N$ and $n'' \in N^-_{w_{(1,2)}}$ and since $\lambda(n'h_{3,1}(b_2)w_{(1,2)}n'') = \lambda(h_{3,1}(b_2))\lambda(w_{(1,2)}n'')$ we can use the assumption on $|b_2|$ and Case 2 to conclude that $\lambda(wn) \le 1$.

Case 5. $w = w_{(1,3,2)}$.
This case is similar to Case 4 and is omitted.

Case 6. $w = w_{(1,3)} = w_0$.
In this case $N_w^- = N$. We first prove two special cases. We show that $\lambda(wx_{1,3}(b_1)) \le 1$ and $\lambda(wx_{1,3}(b_1)x_{2,3}(b_2)) \le 1$. The proof is similar to Case 4 and is omitted. The general case follows the same pattern. Let $wn = wx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3)$. If $|b_3| \le 1$ then $x_{1,2}(b_3) \in K$ and $\lambda(wn) = \lambda(wx_{1,3}(b_1)x_{2,3}(b_2)) \le 1$ by the special case above. Assume $|b_3| > 1$. Let $g' = wnx_{2,1}(-b_3^{-1})$. Then $\lambda(wn) = \lambda(g')$. We can write $g' = n'h_{3,2}(b_3)w_{(1,2)}n''$ for some $n' \in N, n'' \in N^-_{w_{(1,2,3)}}$. Hence $\lambda(g') = \lambda(h_{3,2}(b_3))\lambda(w_{(1,3,2)}n'') \le 1$ by Case 4 and by the condition on $|b_3|$. $\square$

4. Spaces of Whittaker functions.

In this section we define a subspace of the space of Whittaker functions on G and prove some properties of this subspace. In particular we prove Theorem 4.6 which asserts that certain functions on unipotent subgroups are compactly supported. This is one of our main theorems in this paper.

4.1. Whittaker functions. Let ψ_k be a character of k and assume that ψ_k is identically one on O and nontrivial on P^{-1}. For a unipotent matrix $n \in N$ we set

$$\psi(n) = \psi_k(n_{1,2} + n_{2,3}) \tag{4.1}$$

where $n_{i,j}$ are the entries of n. Then

$$\psi(x_{1,2}(b_1)x_{2,3}(b_2)x_{1,3}(b_3)) = \psi_k(b_1 + b_2).$$

We let $\mathcal{W} = \mathcal{W}(G, \psi)$ be the set of functions $W : G \to \mathcal{C}$ such that W is smooth on the right and

$$W(ng) = \psi(n)W(g), \quad n \in N, g \in G.$$

Examples of such functions are Whittaker functions associated with generic representations of G. Other examples are given by projecting compactly supported and locally constant functions to this space as follows:

$$W_f(g) = \int_N f(ng)\psi^{-1}(n)dn, \quad f \in C_c^\infty(G).$$

We shall study the space of such functions $\{W_f : f \in C_c^\infty(G)\}$ in Section 7.

For every $|\alpha| \in |X|$ we extend $|\alpha|$ to G as in (3.1).

For $w \in \mathbf{W}$ we set

$$S^0(w) = S(ww_0) \tag{4.2}$$

where $S(ww_0)$ is defined in (2.12).

Definition 4.1. Let $\mathcal{W} = \mathcal{W}(G, \psi)$ be the space of Whittaker functions defined above. We define $\mathcal{W}^0 = \mathcal{W}^0(G, \psi) \subset \mathcal{W}$ to be the set of functions $W \in \mathcal{W}$ such that for every $w \in \mathbf{W}$ and every $\alpha \in S^0(w)$ there exist positive constants $D_\alpha < E_\alpha$ such that if $g \in BwB$ then

$$W(g) \neq 0 \implies D_\alpha < |\alpha|(g) < E_\alpha, \;\; \alpha \in S^0(w). \tag{4.3}$$

In other words, $W \in \mathcal{W}^0$ if for each $w \in \mathbb{W}$ and each $\alpha \in S^0(w)$ the support of W in BwB has bounded image under α.

Remark 4.2. The condition $|\alpha|(g) < E_\alpha$, $\alpha \in S^0(w)$ that appears in (4.3) is redundant since by [**J-PS-S**] the support of W is contained in the set $\{g : |\alpha|(g) < K, \alpha \in \Delta\}$ for some positive number K. Moreover, if W is a Whittaker function in the Whittaker model of a supercuspidal representation of G then it follows from [**J-PS-S**] that W is compactly supported mod NZ.

Hence it follows that W satisfies condition (4.3) for every $\alpha \in \Delta$ and every $g \in G$ and in particular $W \in \mathcal{W}^0$.

Definition 4.3. Let $\alpha \in \Delta$. we define the sets

$$X_{C_1,C_2}(\alpha) = \{g \in G \mid C_1 < |\alpha|(g) < C_2\},$$
$$A_{C_1,C_2}(\alpha) = \{a \in A \mid C_1 < |\alpha|(a) < C_2\}$$

and the sets

$$X_{C_1,C_2} = \bigcap_{\alpha\in\Delta} X_{C_1,C_2}(\alpha), \quad A_{C_1,C_2} = \bigcap_{\alpha\in\Delta} A_{C_1,C_2}(\alpha).$$

Lemma 4.4. *Let $\alpha \in \Delta$, $C_1 < C_2$ positive numbers and R a compact set in G. Then:*

(a) *There exist constants $C_1' < C_2'$ such that*
$$X_{C_1,C_2}(\alpha)R \subset X_{C_1',C_2'}(\alpha).$$

(b) *Let Y be a subset of G and assume that for every $y \in Y$ there exist $r \in R$ such that $yr \in X_{C_1,C_2}(\alpha)$. Then there exist positive constants $C_1' < C_2'$ such that $Y \subset X_{C_1',C_2'}(\alpha)$.*

Proof. (a) We can write $X_{C_1,C_2}(\alpha) = NA_{C_1,C_2}(\alpha)K$. It is clear that
$$|\alpha|(X_{C_1,C_2}R) = |\alpha|(A_{C_1,C_2}(\alpha))|\alpha|(KR).$$
Since KR is a a compact set in G and $|\alpha|$ is continuous the result follows.

(b) Let $y \in Y$ and let $y = n_0a_0k_0$ be an Iwasawa decomposition for y. If $r \in R$ then $|\alpha|(yr) = |\alpha|(y)||\alpha|(k_0r)$. Since KR is a compact set, there exist positive constants $D_1 < D_2$ such that $D_1 < |\alpha(k_0r)| < D_2$ for all $k_0 \in K$ and $r \in R$. By our assumption, there exists $r_0 \in R$ such that $C_1 < |\alpha|(yr_0) < C_2$. Hence $C_1/D_2 < |\alpha|(y) < C_2/D_1$ and we can choose $C_1' = C_1/D_2$ and $C_2' = C_2/D_1$. □

Corollary 4.5.

(a) *The set $\mathcal{W}^0$ is invariant under right translations by B, i.e., if $W \in \mathcal{W}^0$ then for every $b \in B$, $W_b \in \mathcal{W}^0$ where $W_b(g) = W(gb)$.*

(b) *W^0 is invariant under right integration by compact open subset of closed subgroups of B, i.e., if H is a closed subgroup of B and $X \subset H$ is open and compact in H then for every $W \in \mathcal{W}^0$, $W_X \in \mathcal{W}^0$ where $W_X(g) = \int_X W(gh)dh$.*

Proof. (a) We take R to be the Singleton, $R = \{b^{-1}\}$, where $b \in B$. By Lemma 4.4, $X_{C_1,C_2}(\alpha)b^{-1} \subset X_{C_1',C_2'}(\alpha)$. Thus if W restricted to the set BwB is supported on $X_{C_1,C_2}(\alpha) \cap BwB$ then W_b restricted to BwB will be supported on the set $X_{C_1',C_2'}(\alpha) \cap BwB$.

(b) Since W is smooth on the right, W_X is a finite linear combination of W_{b_i} for some $b_i \in B$, hence (b) follows from (a). □

For each $w \in W$ we let $\mathcal{B}(S^0(w))$ be the basis of $|V|$ defined in (2.10) and let $\mathcal{B}^*(S^0(w))$ be the dual basis (up to scalars) that we fixed in (2.11). Let M be a positive constant. We define the multiplicative cone $A^M(w) \subset A$ to be

$$A^M(w) = \{a \in A : |\beta|(a) < M, \text{ for all } \beta \in \mathcal{B}^*(S^0(w))\}.$$

Our first main theorem of this paper is the following:

Theorem 4.6. *Let $W \in \mathcal{W}^0$ and M a positive constant. Then the function*

$$(a, n) \mapsto W(awn)$$

defined on the set $A^M(w) \times N_w^-$ is compactly supported in N_w^-. That is, if $W(awn) \neq 0$ and $a \in A^M(w), n \in N_w^-$ then n is in some compact set independent of a.

Note that if $w = w_0$ then $S^0(w) = \emptyset$ hence $\mathcal{B}(S^0(w)) = \Delta^*$ and $\mathcal{B}^*(S^0(w)) = \Delta$. It follows that $A^M(w) = A^M$ as defined in (1.2). Since $N_w^- = N$ in that case, Theorem 1.1 follows from the above theorem.

Proof. We prove the theorem for $w = e$, $w = w_{(1,2)}$, $w = w_{(1,2,3)}$ and $w = w_{(1,3)} = w_0$. The case $w = w_{(2,3)}$ is similar to the case $w = w_{(1,2)}$ and the case $w = w_{(1,3,2)}$ is similar to the case $w = w_{(1,2,3)}$, hence they will be omitted.

Case 1. $w = e$.
In this case, $N_w^- = \{e\}$ and there is nothing to prove.

Case 2. $w = w_{(1,2)}$.
In this case

$$w = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and $N_w^- = N_{1,2}$. We need to prove that $W(awx_{1,2}(b)) \neq 0$ implies that there is a constant C independent of $a \in A^M(w)$ such that $|b| < C$. Since W is smooth on the right, there exist a constant C_1 such that $W(gx_{2,1}(-b^{-1})) = W(g)$ for all $|b| > C_1$ and all $g \in G$. Hence, if $|b| > C_1$ it follows from (2.6) that

$$W(awx_{1,2}(b)x_{2,1}(-b^{-1})) = W(\widetilde{n}ah_{2,1}(b)) = \psi(\widetilde{n})W(ah_{2,1}(b))$$

for some $\widetilde{n} \in N$. Hence $W(awx_{1,2}(b)) \neq 0$ implies that $W(ah_{2,1}(b)) \neq 0$. We have that $S^0(e) = \Delta = \{|\alpha|_{1,2}, |\alpha|_{2,3}\}$. Since $W(ah_{1,2}(b)) \neq 0$ and since $W \in \mathcal{W}^0$ it follows that both $|\alpha|_{1,2}$ and $|\alpha|_{2,3}$ are bounded on $ah_{2,1}(b)$. Hence $|\alpha|_{1,3} = |\alpha|_{1,2} + |\alpha|_{2,3}$ is bounded on $ah_{2,1}(b)$. Write $a = d(a_1, a_2, a_3)$. Then $|\alpha|_{1,3}(ah_{2,1}(b)) = \frac{|a_1|}{|ba_3|}$. Hence there exist positive constants E_1 and E_2 such that

$$E_1 < \frac{|a_1|}{|ba_3|} < E_2. \tag{4.4}$$

On the other hand $S^0(w_{1,2}) = \Delta$ and $B^*(S^0(w_{1,2})) = \Delta^* = \{\lambda_1, \lambda_2\}$ (see (2.11)). Since $a \in A^M(w)$, we have that $\lambda_1(a) < M$ and $\lambda_2(a) < M$. Since $|\alpha|_{1,3} = \frac{1}{3}(\lambda_1 + \lambda_2)$ we have that $|\alpha|_{1,3}(a) = \frac{|a_1|}{|a_3|} < M^{2/3}$. Combining this with (4.4) we get that $|b| < E_1 M^{2/3}$.

Case 3. $w = w_{2,3}$.
The proof is similar to Case 2.

Case 4. $w = w_{(1,2,3)}$.
In this case

$$w = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

and $N_w^- = N_{2,3}N_{1,3}$. We need to prove that $W(awx_{1,3}(b_1)x_{2,3}(b_2)) \neq 0$ implies that there is a constant C independent of $a \in A^M(w)$ such that $|b_1| < C$ and $|b_2| < C$.

We first prove a special case. Assume that $W(awx_{1,3}(b_1)) \neq 0$ and $a \in A^M(w)$. If $|b_1|$ is large enough then we have

$$W(awx_{1,3}(b_1)) = W(awx_{1,3}(b_1)x_{3,1}(-b_1^{-1})) = \psi(\widetilde{n})W(ah_{2,1}(b_1)w_{(2,3)})$$

for some $\widetilde{n} \in N$. Since $S^0(w_{(2,3)}) = \Delta$, $w_{(2,3)} \in K$ and $W \in \mathcal{W}^0$ it follows that $|\alpha|_{1,2}$ and $|\alpha|_{2,3}$ are bounded on $ah_{2,1}(b_1)$. In particular, $|\alpha|_{1,3}(ah_{2,1}(b_1)) = (|\alpha|_{1,2} + |\alpha|_{2,3})(ah_{2,1}(b_1)) = |a_1 a_3^{-1} b_1^{-1}|$ is bounded from below.

On the other hand, $S^0(w_{1,2,3}) = \{|\alpha|_{2,3}\} = \{|\alpha|_{0,1,-1}\}$ hence $\mathcal{B}(S^0(w_{(1,2,3)})) = \{|\alpha|_{2,-1,-1}, |\alpha|_{0,1,-1}\}$ and $\mathcal{B}^*(S^0(w_{1,2,3})) = \{|\alpha|_{2,-1,-1}, |\alpha|_{0,1,-1}\}$ (see (2.11)). Since $a \in A^M(w)$ we have that $|\alpha|_{1,3}(a) = (\frac{1}{2}(|\alpha|_{2,-1,-1} + |\alpha|_{0,1,-1}))(a) = |a_1 a_3^{-1}| < M$ hence it follows that $|b_1|$ is bounded from above which is what we need.

For the general case, assume $W(awx_{1,3}(b_1)x_{2,3}(b_2)) \neq 0$ and $a \in A^M(w)$. If $|b_2|$ is large enough then we can write

$$\begin{aligned} W(awx_{1,3}(b_1)x_{2,3}(b_2)) &= W(awx_{1,3}(b_1)x_{2,3}(b_2)x_{3,2}(-b^{-1})) \\ &= \psi(\widetilde{n})W(ah_{3,1}(b_2)w_{(1,2)}x_{1,2}(-b_1 b_2^{-1})) \end{aligned}$$

for some $\widetilde{n} \in N$. This lands us in Case 2! To use Case 2, we need to show that if $a \in A_M(w_{(1,2,3)})$ and $|b_2|$ is large then $ah_{3,1}(b_2) \in A^{M_1}(w_{(1,2)})$ for some positive constant M_1. Assume $|b_2| > D$. Then

$$\begin{aligned} |\alpha|_{2,-1,-1}(ah_{3,1}(b_2)) &= |\alpha|_{2,-1,-1}(a)|\alpha|_{2,-1,-1}(h_{3,1}(b_2)) \\ &= |\alpha|_{2,-1,-1}(a)||b_2^{-3}| < MD^{-3}. \end{aligned}$$

Also

$$|\alpha|_{1,1,-2}(ah_{3,1}(b_2)) = |\alpha|_{1,1,-2}(a)|\alpha|_{1,1,-2}(h_{3,1}(b_2))$$
$$= \left(\frac{1}{2}(|\alpha|_{2,-1,-1} + 3|\alpha|_{0,1,-1})\right)(a)|b_2^{-3}| < M^2D^{-3}.$$

Hence we can choose $M_1 = \max\{MD^{-3}, M^2D^{-3}\}$.

It follows that we can now use Case 2 to conclude that

$$W(ah_{3,1}(b_2)w_{(1,2)}x_{1,2}(-b_1b_2^{-1})) \neq 0$$

with the above conditions on a and b_2 implies that $|-b_1b_2^{-1}|$ is bounded.

Now $S^0(w_{1,2}) = \Delta$. Since $W \in \mathcal{W}^0$ it follows that $|\alpha|_{2,3}$ is bounded on $ah_{3,1}(b_2)w_{(1,2)}x_{1,2}(-b_1b_2^{-1})$. (Here b_1 and b_2 and a are allowed to change but are subject to the conditions $|b_2| > C$, $a \in A^M(w)$ and $W(ah_{3,1}(b_2)w_{(1,2)}x_{1,2}(-b_1b_2^{-1})) \neq 0$.) Since $|-b_1b_2^{-1}|$ is bounded it follows from Lemma 4.4 (b) that $|\alpha|_{2,3}$ is bounded on $ah_{3,1}(b_2)$. In particular, $|\alpha|_{2,3}(ah_{3,1}(b_2)) = |a_2a_3^{-1}b_2^{-1}|$ is bounded from below. Since $a \in A^M(w)$ we get that $|\alpha|_{2,3}(a) = |a_2a_3^{-1}| < M$ hence we get that $|b_2|$ is bounded from above (independent of b_1 or $a \in A^M(w)$).

So we can assume that $|b_2| \leq C_1$ for some positive constant C_1. Since W is smooth on the right it follows that the space spanned by the functions $\{\rho(x_{2,3}(b_2))W \ : \ |b_2| \leq C_1\}$ is finite dimensional. Here $(\rho(h)W)(g) = W(gh)$. Let $W_1, \ldots, W_l$ be a basis for this space. It follows from Corollary 4.5 (a) that W_i, $i = 1, \ldots, l$ are all in $\mathcal{W}^0$. By the special case above it follows that if $a \in A^M(w)$ and $W_i(awx_{1,3}(b_1)) \neq 0$ the there exist a constant D_i independent of a such that $|b_1| < D_i$. Let D be the maximum of the D_is. We write

$$W(awx_{1,3}(b_1)x_{2,3}(b_2)) = (\rho(x_{2,3}(b_2))W)(awx_{1,3}(b_1))$$
$$= \sum_i \lambda_i W_i(awx_{1,3}(b_1))$$

where λ_i is a scalar depending on b_2. Since $W(awx_{1,3}(b_1)x_{2,3}(b_2)) \neq 0$, it follows that there exist i such that $W_i(awx_{1,3}(b_1)) \neq 0$ hence $|b_1| < D$. We notice that D is independent of $|b_2| \leq C_1$.

Case 5. $w = w_{(1,3,2)}$.

This case is proved in the same way as Case 4.

Case 6. $w = w_{(1,3)} = w_0$.

This case is similar to Case 4. We will outline the proof:

In this case $N_w^- = N$. We need to prove that

$$W(awx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3)) \neq 0$$

implies that there is a constant C independent of $a \in A^M(w)$ such that $|b_1| < C$, $|b_2| < C$ and $|b_3| < C$. We first prove two special cases. In the first case we show that $W(awx_{1,3}(b_1)) \neq 0$ implies that there is a constant C independent of $a \in A^M(w)$ such that $|b_1| < C$. In the second case we show that $W(awx_{1,3}(b_1)x_{2,3}(b_2)) \neq 0$ implies that there is a constant C independent of $a \in A^M(w)$ such that $|b_1| < C$ and $|b_2| < C$. The proof follows the same lines as in Case 4 and is omitted.

For the general case we assume that $|b_3|$ is large. Then

$$awx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3)x_{2,1}(-b_3^{-1}) = n'ah_{3,2}(b_3)w_{(1,2,3)}n''$$

for some $n' \in N$ and $n'' \in N^-_{w_{(1,2,3)}}$. Hence

$$\begin{aligned} W(&awx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3)) \\ &= W(awx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3)x_{2,1}(-b_3^{-1})) \\ &= \psi(n')W(ah_{3,2}(b_3)w_{(1,3,2)}n''). \end{aligned} \tag{4.5}$$

Here we landed in Case 4. The proof continues as in the general step of Case 4. $\square$

5. Projection into $\mathcal{W}^0(G, \psi)$.

In this section we shall show that every $W \in \mathcal{W}(G, \psi)$ can be projected into $\mathcal{W}^0(G, \psi)$ by integrating it on a compact unipotent group versus a character of that group. We start with some preliminary results about Howe vectors.

5.1. Howe vectors. For a positive integer m we denote by K_m the congruence subgroup of K given by $K_m = I_3 + M_3(P^m)$. We let $A_m = A \cap K_m$. Let

$$d = \begin{pmatrix} 1 & & \\ & \varpi^2 & \\ & & \varpi^4 \end{pmatrix}.$$

Let $J_m = d^m K_m d^{-m}$. Then J_m is given by

$$J_m = I_3 + \begin{pmatrix} P^m & P^{-m} & P^{-3m} \\ P^{3m} & P^m & P^{-m} \\ P^{5m} & P^{3m} & P^m \end{pmatrix}. \tag{5.1}$$

Notice that J_m is expanding above the main diagonal and shrinking on and below the main diagonal. Let

$$N_m = N \cap J_m. \tag{5.2}$$

Let $\overline{N}_m = \overline{N} \cap J_m$ and $\overline{B}_m = \overline{B} \bigcap J_m$. It is easy to see that

$$J_m = \overline{N}_m A_m N_m = \overline{B}_m N_m.$$

We fix a character ψ_k on k as Section 4. In particular $\psi_k(O) = 1$ and $\psi_k(P^{-1}) \neq 1$. Let ψ be a character of N obtained from ψ as in (4.1). For $m \geq 1$ we define a character ψ_m on J_m by

$$\psi_m(j) = \psi(n_j)$$

where $j = \bar{b}_j n_j$, $\bar{b}_j \in \overline{B}_m$, $n_j \in N_m$ is the unique decomposition of j. It is easy to see that ψ_m is a character on J_m. For each $W \in \mathcal{W}(G, \psi)$ we define $W_m = W_{N_m,\psi}$ by

$$W_m(g) = W_{N_m,\psi}(g) = \int_{N_m} W(gn)\psi^{-1}(n)dn. \tag{5.3}$$

Since $N_{m+1} \supset N_m$ it is a simple application of Fubini to show that if $m \geq k$ then

$$W_m(g) = \mathrm{vol}(N_k)^{-1} \int_{N_m} W_k(gn)\psi^{-1}(n)dn. \tag{5.4}$$

For $g_1 \in G$ we let $(\rho(g_1)W)(g) = W(gg_1)$.

Lemma 5.1. *Let M be such that $\rho(K_M)W = W$ and let m be an integer such that $m > 3M$. Then*

$$\rho(j)W_m = \psi_m(j)W_m, \quad j \in J_m. \tag{5.5}$$

Proof. Define

$$\widetilde{W}_m = \mathrm{vol}(\overline{B}_m)^{-1} \int_{J_m} \psi_m^{-1}(j)\rho(j)W_m dj.$$

It is clear that $\widetilde{W}_m$ satisfies (5.5). On the other hand we have

$$\begin{aligned} &\mathrm{vol}(\overline{B}_m)^{-1} \int_{J_m} \psi_m^{-1}(j)\rho(j)W_m dj \\ &= \mathrm{vol}(\overline{B}_m)^{-1} \int_{N_m} \int_{\overline{B}_m} \psi^{-1}(n)\rho(n)\rho(\bar{b})W \, dn \, db. \end{aligned}$$

Looking at (5.1) it is easy to see that $\overline{B}_m \subseteq K_M$, hence $\bar{b} \in \overline{B}_m$ fixes W, and we get that $\widetilde{W}_m = W_m$. □

Formulating Lemma 5.1 for functions we get that

$$W_m(gj) = \psi_m(j)W_m(g) \ \text{ for all } \ g \in G, j \in J_m. \tag{5.6}$$

We call a vector W in a representation space of G satisfying (5.5) (or (5.6)) a Howe vector. The above Lemma shows that if the representation space affords a nontrivial Whittaker functional then nonzero Howe vectors exist. This property and some uniqueness properties of Howe vectors for irreducible admissible representations of $\mathrm{GL}_n(k)$ were established in [**H**]. We now continue to study the behavior of Whittaker functions satisfying (5.6).

Lemma 5.2. *Let $m > 0$ and assume that $W \in \mathcal{W}(G, \psi)$ satisfies (5.6). Let $a \in A$. Then $W(a) \neq 0$ implies that $\alpha_{1,2}(a) \in 1 + P^m$ and $\alpha_{2,3}(a) \in 1 + P^m$. In particular, $|\alpha|_{1,2}$ and $|\alpha|_{2,3}$ are bounded on a.*

Proof. Assume $|b| \leq q^m$. Then

$$\psi(b)W(a) = W(ax_{1,2}(b)) = W(x_{1,2}(\alpha_{1,2}(a)b)a) = \psi(\alpha_{1,2}(a)b)W(a).$$

Since $W(a) \neq 0$ then $\psi(b) = \psi(\alpha_{1,2}(a)b)$ for every b such that $|b| \leq q^m$ hence using that ψ has conductor O we get that $\alpha_{1,2}(a) \in 1 + P^m$. A similar proof works for $\alpha_{2,3}$. $\square$

Let χ_X be the characteristic function of the set X.

Corollary 5.3. *Assume that $W \in \mathcal{W}(G, \psi)$ satisfies* (5.6). *Then*

$$W\begin{pmatrix} a_1 & & \\ & a_2 & \\ & & 1 \end{pmatrix} = \chi_{1+P^m}(a_1)\chi_{1+P^m}(a_2).$$

Let $N_2 \subset \mathrm{GL}(2, k)$ be the subgroup of upper unipotent matrices and let

$$\overline{B}_{2,m} = \left\{ \begin{pmatrix} x_1 & 0 \\ x_2 & x_3 \end{pmatrix} \ : \ x_1, x_3 \in 1 + P^m, x_2 \in P^{3m} \right\}.$$

The following Lemma is not needed for the proof of the main results in this paper. It is needed for our result on Bessel distributions in [**Ba2**].

Lemma 5.4. *Assume that $W \in \mathcal{W}(G, \psi)$ satisfies* (5.6). *Let $h \in \mathrm{GL}(2, k)$. Then*

$$W\begin{pmatrix} h & \\ & 1 \end{pmatrix} = 0$$

if $h \notin N_2\overline{B}_{2,m}$. If $h = n\bar{b}$ for $n \in N_2$ and $\bar{b} \in \overline{B}_{2,m}$ then

$$W\begin{pmatrix} h & \\ & 1 \end{pmatrix} = \psi(n)W(e)$$

where ψ is the natural restriction of ψ to N_2.

Proof. Let

$$h = \begin{pmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{pmatrix}$$

and let

$$\begin{pmatrix} I & y \\ & 1 \end{pmatrix} \in J_m$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

and the above matrix being in J_m is equivalent to $y_1 \in P^{-3m}$ and $y_2 \in P^{-m}$. We have

$$\begin{aligned}\psi(y_2)W\begin{pmatrix} h & \\ & 1\end{pmatrix} &= W\left(\begin{pmatrix} h & \\ & 1\end{pmatrix}\begin{pmatrix} I & y \\ & 1\end{pmatrix}\right)\\ &= W\left(\begin{pmatrix} I & hy \\ & 1\end{pmatrix}\begin{pmatrix} h & \\ & 1\end{pmatrix}\right)\\ &= \psi(h_{2,1}y_1 + h_{2,2}y_2)W\begin{pmatrix} h & \\ & 1\end{pmatrix}.\end{aligned}$$

Assume that

$$W\begin{pmatrix} h & \\ & 1\end{pmatrix} \neq 0.$$

Then we have $\psi(h_{2,1}y_1 + (h_{2,2} - 1)y_2) = 1$ for all y_1 and y_2 as above hence by the assumptions on the conductor of ψ we have that $h_{2,1} \in P^{3m}$ and $h_{2,2} \in 1 + P^m$. It follows that we can write

$$h = \begin{pmatrix} y_1 & y_2 \\ 0 & y_3\end{pmatrix}\begin{pmatrix} 1 & 0 \\ x & 1\end{pmatrix}$$

with $x \in P^{3m}$. Using (5.6) and Corollary 5.3 we get that $h \in N_2\overline{B}_{2,m}$ hence we proved the first statement. Using the equivariance of W on the left by N and on the right by J_m we get the second statement. □

Lemma 5.5. *Assume that $W \in \mathcal{W}(G, \psi)$ satisfies* (5.6)*. Then $W(aw_{(1,2)}) = 0$ and $W(aw_{(2,3)}) = 0$ for every $a \in A$.*

Proof. Let $b \in P^{-m}$ be such that $\psi(b) \neq 1$. Then

$$\begin{aligned}\psi(b)W(aw_{(1,2)}) &= W(aw_{(1,2)}x_{2,3}(b))\\ &= W(x_{1,3}(\alpha_{1,3}(a)b)aw_{(1,2)}) = W(aw_{(1,2)}).\end{aligned}$$

Since $\psi(b) \neq 0$ we get that $W(aw_{(1,2)}) = 0$. A similar proof will show that $W(aw_{2,3}) = 0$. □

Lemma 5.6. *Assume that $W \in \mathcal{W}(G, \psi)$ satisfies* (5.6)*. Then $W(aw_{(1,2,3)}) \neq 0$ implies that $\alpha_{2,3}(a) \in 1 + P^m$ and in particular $|\alpha|_{2,3}$ is bounded on a. Also, $W(aw_{(1,3,2)}) \neq 0$ implies that $\alpha_{1,2}(a) \in 1 + P^m$ and in particular $|\alpha|_{1,2}$ is bounded on a.*

Proof. Let $b \in P^{-m}$. Then

$$\begin{aligned}\psi(b)W(aw_{(1,2,3)}) &= W(aw_{(1,2,3)}x_{1,2}(b))\\ &= W(x_{2,3}(\alpha_{2,3}(a)b)aw_{(1,2,3)})\\ &= \psi(\alpha_{2,3}(a)b)W(aw_{(1,2,3)}).\end{aligned}$$

Hence, $W(aw_{(1,2,3)}) \neq 0$ implies that $\psi(b) = \psi(\alpha_{2,3}(a)b)$ for all $b \in P^{-m}$ hence $\alpha_{2,3}(a) \in 1+P^m$. A similar proof works for the case where $W(aw_{(1,3,2)}) \neq 0$. □

We now prove the second main theorem of this paper. It was stated in the introduction as Theorem 1.3. Let $\mathcal{W}^0(G,\psi)$ be the subspace of $\mathcal{W}(G,\psi)$ introduced in Definition 4.1.

Theorem 5.7. *Let $W \in \mathcal{W}(G,\psi)$. Then there exist a positive integer M such that $W_m = W_{N_m,\psi} \in \mathcal{W}^0(G,\psi)$ for every $m \geq M$.*

Proof. We need to show that there exist M such that for every fixed $m \geq M$ and every $w \in \mathbb{W}$, the support of W_m in BwB has bounded image under every $|\alpha| \in S^0(w)$. It is enough to show this for each individual cell and then take the maximum of these Ms that we obtain.

We now show that given $w \in \mathbb{W}$ there exist an integer M such that if $m \geq M$ and if $W_m(g) \neq 0$ then $|\alpha|(g)$ is in a fixed bounded set in $\mathbf{R}^*$ for every $\alpha \in S^0(w)$.

Case 1. $w = e$.

In that case $S^0(w) = \{|\alpha|_{1,2}, |\alpha|_{2,3}\}$. Let M be such that W_m is a Howe vector (that is satisfies (5.6)) for $m \geq M$. We need to show that if $g \in BwB = B$ and $W_m(g) \neq 0$ then both $|\alpha|_{1,2}$ and $|\alpha|_{2,3}$ are bounded on g. Since $g \in B$ we can write $g = na$ for $n \in N$ and $a \in A$. We have that $|\alpha|(g) = |\alpha|(a)$ for every $|\alpha| \in |X|$. Also, $W_m(g) = \psi(n)W_m(a)$ hence $W_m(g) \neq 0$ implies that $W_m(a) \neq 0$. Now the theorem follows from Lemma 5.2.

Case 2. $w = w_{(1,2)}$.

In that case $S^0(w) = \{|\alpha|_{1,2}, |\alpha|_{2,3}\}$. Let M_1 be such that W_m is a Howe vector for $m \geq M_1$ and let $M = 3M_1$. Let $m \geq M$. Let $g \in BwB$ and assume $W_m(g) \neq 0$. We can write $g = naw_{(1,2)}x_{1,2}(b)$ for some $n \in N$, $a \in A$ and $b \in k$. To make the proof more clear we shall consider three cases although two may suffice. The first case is when $|b| \geq q^{3m}$. In that case $x_{2,1}(-b^{-1}) \in J_m$ hence by (5.6) we have

$$W_m(g) = W(gx_{2,1}(-b^{-1})) = W(\widetilde{g})$$

where $\widetilde{g} = gx_{2,1}(-b^{-1}) \in B$. By Case 1 it follows that $W(\widetilde{g}) \neq 0$ implies that $|\alpha|_{1,2}(\widetilde{g})$ and $|\alpha|_{2,3}(\widetilde{g})$ are in a fixed compact set. Hence if we let $R = \{x_{2,1}(-b^{-1}) \;:\; |b| \geq q^{3m}\} \cup \{e\}$ then R is a compact set in G and the above argument implies that for each g of the above form such that $|b| \geq q^{3m}$ and such that $W_m(g) \neq 0$ there exist $r = r_g \in R$ such that $|\alpha|_{1,2}(gr)$ and $|\alpha|_{2,3}(gr)$ are in fixed compact set of $\mathbf{R}^*$. Hence, by Lemma 4.4 (b) we get that $|\alpha|_{1,2}(g)$ and $|\alpha|_{2,3}(g)$ are in compact sets independent of g, (depending only on m and W).

The second case is when $q^m < |b| < q^{3m}$. By (5.4) we can write

$$W_m(g) = \frac{1}{\mathrm{vol}(N_{M_1})} \int_{N_m} W_{M_1}(gn)\psi^{-1}(n)dn.$$

Hence, if $W_m(g) \neq 0$ there exist $n' \in N_m$ such that $W_{M_1}(gn') \neq 0$. We have that $gn' = n''aw_{(1,2)}x_{1,2}(\widetilde{b})$ for some $n'' \in N$ and $\widetilde{b} \in k$ such that

$|\widetilde{b}| = |b|$. Since $|b| > q^m \geq q^M = q^{3M_1}$ it follows that $x_{2,1}(-\widetilde{b}^{-1}) \in J_{M_1}$ and we can use the argument above to conclude that $|\alpha|_{1,2}(gn'x_{2,1}(-\widetilde{b}^{-1}))$ and $|\alpha|_{2,3}(gn'x_{2,1}(-\widetilde{b}^{-1}))$ are in fixed compact sets in $\mathbf{R}^*$ independent of g. Let $Q = \{x_{2,1}(c) : \ |c| \leq q^{-m}\}$. Then $R = N_m Q$ is a compact set in G and we can use Lemma 4.4 (b) to conclude that $|\alpha|_{1,2}(g)$ and $|\alpha|_{2,3}(g)$ are in fixed compact sets in $\mathbf{R}^*$ for all such g.

The last case is when $|b| \leq q^m$. In that case $x_{1,2}(b) \in J_m$ and we have

$$W_m(g) = W_m(naw_{(1,2)}x_{1,2}(b)) = \psi(n)\psi_k(b)W_m(aw_{(1,2)}) = 0$$

where the last equality is Lemma 5.5.

Case 3. $w = w_{(2,3)}$.

This case is similar to Case 2 and is omitted.

Case 4. $w = w_{(1,2,3)}$.

In this case $S^0(w) = \{|\alpha|_{2,3}\}$. Let M_2 be such that W_m satisfies (5.6) and such that W_m satisfies the support conditions for the theorem on the Bruhat cells covered by Cases 1, 2 and 3 for every $m \geq M_2$. (Since an appropriate M exist for each case we can take M_2 to be the maximum of the three.) Let $M_1 = 3M_2$ and $M = 3M_1 = 9M_2$. We need to show that for a fixed $m \geq M$, $|\alpha|_{2,3}$ is bounded on the set of $g \in BwB$ such that $W_m(g) \neq 0$. We can write

$$g = nawx_{1,3}(b_1)x_{2,3}(b_2) \tag{5.7}$$

where $n \in N$, $a \in A$ and $b_1, b_2 \in k$. We first show that for a fixed $m \geq M_1$, $|\alpha|_{2,3}$ is bounded on the set of g of the form (5.7) satisfying $W_m(g) \neq 0$ and $|b_2| > q^m$. In that case we can write

$$W_m(g) = \frac{1}{\mathrm{vol}(N_{M_2})} \int_{N_m} W_{M_2}(gn)\psi^{-1}(n)dn.$$

Since $W_m(g) \neq 0$ we get that $W_{M_2}(gn') \neq 0$ for some $n' \in N_m$. Writing gn' in the form (5.7) with $\widetilde{b}_2$ replacing b_2 it is easy to see that $|\widetilde{b}_2| = |b_2| > q^m \geq q^{M_1} = q^{3M_2}$. Hence $x_{3,2}(-\widetilde{b}_2^{-1}) \in J_{M_2}$. It follows that $W_{M_2}(gn') = W_{M_2}(gn'x_{3,2}(-\widetilde{b}_2^{-1})) \neq 0$. However, $\widetilde{g} = gn'x_{3,2}(-\widetilde{b}_2^{-1}) \in Bw_{(1,2)}B$ and our assumptions on M_2 imply that $|\alpha|_{2,3}(\widetilde{g})$ is in a fixed compact set in $\mathbf{R}^*$ depending only on M_2. Hence by Lemma 4.4 (b) $|\alpha|_{2,3}(g)$ is in a fixed compact set depending only on m and M_2.

Now assume $m \geq M$, $W_m(g) \neq 0$ and g is of the form (5.7) with $|b_2| \leq q^m$. If $|b_1| \leq q^{3m}$ then $x_{1,3}(b_1)x_{2,3}(b_2) \in J_m$ and we have

$$W_m(g) = \psi(n)\psi_k(b_1)\psi_k(b_2)W(aw_{(1,2,3)}) \neq 0.$$

By Lemma 5.6 we have that $|\alpha|_{2,3}(a)$ is in a fixed compact set in $\mathbf{R}$. But $|\alpha|_{2,3}(a) = |\alpha|_{2,3}(gx_{2,3}(-b_2)x_{1,3}(-b_1)) = |\alpha|_{2,3}(g\widetilde{n})$ for $\widetilde{n} \in N_m$. Hence, by Lemma 4.4 (b) we get that $|\alpha|_{2,3}(g)$ is in a fixed compact set.

Now assume $|b_1| > q^{3m}$. As before we write

$$W_m(g) = \frac{1}{\mathrm{vol}(N_{M_1})} \int_{N_m} W_{M_1}(gn)\psi^{-1}(n)dn.$$

Again we have that there exist $n' \in N_m$ such that $W_{M_1}(gn') \neq 0$. Write

$$g' = gn' = n''awx_{1,3}(\widetilde{b}_1)x_{2,3}(\widetilde{b}_2)$$

as in (5.7). It is easy to see that $|\widetilde{b}_1| = |b_1| > q^{3m} \geq q^{3M} = q^{9M_1}$. We now separate into two cases. If $|\widetilde{b}_2| > q^{M_1}$ then we are in the case already discussed above hence we have that $|\alpha|_{2,3}(gn')$ is in a fixed compact set hence $|\alpha|_{2,3}(g)$ is in a fixed compact set. Now assume that $|\widetilde{b}_2| \leq q^{M_1}$. Then $x_{2,3}(\widetilde{b}_2) \in J_m$ and

$$W_{M_1}(g') = \psi_k(\widetilde{b}_2)W_{M'}(n''awx_{1,3}(\widetilde{b}_1)) \neq 0.$$

Since $|\widetilde{b}_1| \geq q^{9M_1}$ we have that $x_{3,1}(-\widetilde{b}_1^{-1}) \in J_{M_1}$. Hence

$$W_{M_1}(n''awx_{1,3}(\widetilde{b}_1)) = W_{M_1}(n''awx_{1,3}(\widetilde{b}_1)x_{3,1}(-\widetilde{b}_1^{-1})) \neq 0.$$

However, $g'' = n''awx_{1,3}(\widetilde{b}_1)x_{3,1}(-\widetilde{b}_1^{-1}) \in Bw_{(2,3)}B$ and by our assumptions on M_1, $|\alpha|_{2,3}(g'')$ is in a fixed compact set. Since $g'' = g'x_{3,1}(-\widetilde{b}_1^{-1}) = gn'x_{2,3}(-\widetilde{b}_2)x_{3,1}(-\widetilde{b}_1^{-1})$ we get that $|\alpha|_{2,3}(g)$ is in a fixed compact set.

Case 5. $w = w_{(1,3,2)}$.
This case is similar to Case 4 and is omitted.

Case 6. $w = w_{(1,3)} = w_0$.
In this case, $S^0(w) = \emptyset$ and there is nothing to prove. □

6. Bessel functions.

In this section we attach Bessel functions to generic representations of G. Each Bessel function will be defined on one of three Bruhat cells and will be left and right invariant by (N, ψ). The most significant one is the Bessel function attached to the open Bruhat cell Bw_0B.

Let $w \in \mathbb{W}$. Let N_w^+ and N_w^- be the subgroups of N as defined in (2.14). We define the subtorus A_w to be

$$A_w = \{a \in A : \ \psi(n) = \psi(n^{aw}), \ \text{for all } n \in N_w^+\}. \tag{6.1}$$

Here $n^g = gng^{-1}$. It is easy to see that $A_e = Z(G)$, $A_{(1,2)} = A_{(2,3)} = \emptyset$, $A_{(1,2,3)} = \{d(a_1, a_2, a_3) \in A : \ a_2 = a_3\}$, $A_{(1,3,2)} = \{d(a_1, a_2, a_3) \in A : \ a_1 = a_2\}$ and $A_{(1,3)} = A$.

Let $w = w_{(1,2,3)}$ or $w = w_{(1,3,2)}$. It is easy to check that N_w^+ normalizes N_w^-. If $n^+ \in N_w^+$ then the mapping $u \mapsto n^+u(n^+)^{-1}$ is measure preserving on N_w^- and we also have

$$\psi(n^+u(n^+)^{-1}) = \psi(u), \quad \text{for every } u \in N_w^-. \tag{6.2}$$

Let N_m be the open compact subgroup of N defined in (5.2). Then $N_m = (N_m \cap N_w^+)(N_m \cap N_w^-)$. We let $N_m^+ = N_{m,w}^+ = (N_m \cap N_w^+)$ and $N_m^- = N_{m,w}^- = (N_m \cap N_w^-)$. We let $dn = dn^+dn^-$ be a Haar measure on N where dn^+ and dn^- are Haar measures on N_w^+ and N_w^- respectively.

Proposition 6.1. *Let $w \in \mathbb{W}$ and $g \in NA_wwN_w^-$. Let $W \in \mathcal{W}^0(G,\psi)$ and let $W_m = W_{N_m,\psi}$ be the function defined in* (5.3)*. Then*

$$\int_{N_w^-} W(gn)\psi^{-1}ndn \;=\; \frac{1}{\mathrm{vol}(N_m)} \int_{N_w^-} W_m(gn)\psi^{-1}(n)dn.$$

Proof. By Theorem 4.6 the integrals above converge absolutely. If $w = e$ then $N_w^- = \{e\}$ and there is nothing to prove. If $w = w_{(1,2)}$ or $w = w_{(2,3)}$ then $A_w = \emptyset$ and there is nothing to prove. If $w = w_{(1,3)}$ then $N_w^- = N$ and the proposition follows from a simple use of Fubini's theorem. So we assume $w = w_{(1,2,3)}$ or $w = w_{(1,3,2)}$. We first notice that if $g = n_1awn_2$ where $n_1 \in N$ and $n_2 \in N_w^-$ then

$$\int_{N_w^-} W(gn)\psi^{-1}(n)dn = \psi(n_1)\psi(n_2)\int_{N_w^-} W(awn)\psi^{-1}(n)dn$$

and similarly for the second integral. Hence, using the assumption on g, it is enough to prove the equality for $g = aw$ with $a \in A_w$. In that case we have

$$\begin{aligned}
&\frac{1}{\mathrm{vol}(N_m)} \int_{N_w^-} W_m(awu)\psi^{-1}(u)du \\
&= \frac{1}{\mathrm{vol}(N_m^+)}\frac{1}{\mathrm{vol}(N_m^-)} \int_{N_w^-} \left(\int_{N_m^-}\int_{N_m^+} W(awun^+n^-) \right. \\
&\qquad\qquad \left. \cdot\, \psi^{-1}(u)\psi^{-1}(n^+n^-)dn^+dn^- \right) du.
\end{aligned}$$

It follows from Theorem 4.6 that the function $u \mapsto W(awun^+n^-)$ is supported on a compact set in N_w^- independent of $n^+ \in N_m^+$ and $n^- \in N_m^-$ hence the above integral converges absolutely and we can use Fubini and the change of variables $u \mapsto (n^+)^{-1}un^+$ and (6.2) to get that the above integral equals

$$\frac{1}{\mathrm{vol}(N_m^+)}\frac{1}{\mathrm{vol}(N_m^-)}\int_{N_m^-}\int_{N_m^+}\left(\int_{N_w^-}W(awn^+(n^+)^{-1}un^+n^-)\right.$$
$$\left.\cdot\,\psi^{-1}(u)\psi^{-1}(n^+n^-)du\right)dn^+dn^-$$
$$=\frac{1}{\mathrm{vol}(N_m^+)}\frac{1}{\mathrm{vol}(N_m^-)}\int_{N_m^-}\int_{N_m^+}\left(\int_{N_w^-}W(awn^+un^-)\right.$$
$$\left.\cdot\,\psi^{-1}(u)\psi^{-1}(n^+)\psi^-(n^-)du\right)dn^+dn^-.$$

By (6.1) it follows that if $a \in A_w$ and $n^+ \in N_w^+$ then $W(awn^+un^-) = \psi(n^+)W(awun^-)$. Hence the last integral equals

$$\frac{1}{\mathrm{vol}(N_m^-)}\int_{N_m^-}\left(\int_{N_w^-}W(awun^-)\psi^{-1}(un^-)du\right)dn^-$$
$$=\frac{1}{\mathrm{vol}(N_m^-)}\int_{N_m^-}\big(W(awu)\psi^{-1}(u)du\big)\,dn^-$$
$$=\int_{N_w^-}W(awu)\psi^{-1}(u)du^-.$$

□

Let $w \in \mathbb{W}$ and $g \in NA_wwN_w^-$. Let $W \in \mathcal{W}(G,\psi)$. By Theorem 5.7 there exist M such that $W_m \in \mathcal{W}^0$ for every $m \geq M$. Let $m \geq M$. We define

$$J_{g,w}(W) = \frac{1}{\mathrm{vol}(N_m)}\int_{N_w^-}W_m(gn)\psi^{-1}(n)dn. \tag{6.3}$$

By applying Proposition 6.1 to $F = W_M$ and using that $F_m = \mathrm{vol}(N_M)W_m$ if $m \geq M$ we see that the above integral equals

$$\frac{1}{\mathrm{vol}(N_M)}\int_{N_w^-}W_M(gn)\psi^{-1}(n)dn$$

hence it is independent of m.

Remark 6.2. It follows from Proposition 6.1 that if $W \in \mathcal{W}^0$ then

$$J_{g,w}(W) = \int_{N_w^-}W(gn)\psi^{-1}(n)dn.$$

Remark 6.3. For $w = w_0 = w_{(1,3)}$ we have that $N_w^- = N$. In that case we have that if f is a compactly supported function on N then $\int_N f(n)dn =$

$\lim_{m\to\infty}\int_{N_m} f(n)dn$. Hence

$$\begin{aligned}J_{g,w}(W) &= \frac{1}{\mathrm{vol}(N_M)}\int_N W_M(gn)\psi^{-1}(n)dn\\ &= \frac{1}{\mathrm{vol}(N_M)}\lim_{m\to\infty}\int_{N_m} W_M(gn)\psi^{-1}(n)dn\\ &= \lim_{m\to\infty} W_m(g)\\ &= \lim_{m\to\infty}\int_{N_m} W(gn)\psi^{-1}(n)dn\end{aligned}$$

which can be viewed as a principal valued p-adic integral. (See Definition 1.5.)

Lemma 6.4. *Let $n \in N$. Then $J_{g,w}(\rho(n)W) = \psi(n)J_{g,w}(W)$. Here $(\rho(n)W)(g) = W(gn)$.*

Proof. Let $F = \rho(n)W$. Pick m large enough so that $F_m \in \mathcal{W}^0$ and such that $n \in N_m$. For such m we have that $F_m = \psi(n)W_m$ and the result follows from the definition of $J_{g,w}$. □

6.1. Bessel functions. We now return to the setting of Section 1.1. Let π be an irreducible admissible representation of G on a complex vector space V. Let L be a nonzero ψ Whittaker functional on V. Let $\mathcal{W}(\pi,\psi)$ be the space of functions

$$W_v(g) = L(\pi(g)v)$$

for all $v \in V$. It is easy to see that $W_v \in \mathcal{W}(G,\psi)$ and that the space $\mathcal{W}(\pi,\psi) = \{W_v : v \in V\}$ is a subspace of $\mathcal{W}(G,\psi)$ which is isomorphic to (π, V). G acts on $\mathcal{W}(G,\psi)$ by right translations (that is, by ρ). Let $w \in W$ and $g \in NA_w wN_w^-$. It follows from Lemma 6.4 that the restriction of the functional $J_{g,w}$ defined in (6.3) to $\mathcal{W}(\pi,\psi)$ is a Whittaker functional. From the uniqueness of the Whittaker functional it follows that there exist a scalar $j_{\pi,\psi,w}(g)$ such that

$$J_{g,w}(W) = j_{\pi,\psi,w}(g)W(e). \tag{6.4}$$

If $w = w_0 = w_{(1,3)}$ then we let $j_{\pi,\psi}(g) = j_{\pi,\psi,w_0}(g)$ and we call $j_{\pi,\psi}(g)$ the Bessel function of π.

Remark 6.5. The Bessel functions defined above are independent of the choice of initial Whittaker functional L. They do depend on the choice of Haar measure for N.

We let $\mathcal{W}^0(\pi,\psi) = \mathcal{W}(\pi,\psi) \cap \mathcal{W}^0(G,\psi)$.

Remark 6.6. Since L is assumed to be nonzero there exist $W \in \mathcal{W}(\pi,\psi)$ such that $W(e) \neq 0$. Since $W_m(e) = \mathrm{vol}(N_m)W(e)$ it follows from Theorem 5.7 that there exist $W \in \mathcal{W}^0(\pi,\psi)$ such that $W(e) \neq 0$. Hence,

$\mathcal{W}^0(\pi,\psi) \neq \{0\}$. (This is Corollary 1.4.) If we choose $W \in \mathcal{W}^0(\pi,\psi)$ such that $W(e) = 1$ we get that $J_{g,w}(W) = j_{\pi,\psi,w}(g)$. By Remark 6.2 it follows that for such W we have

$$j_{\pi,\psi,w}(g) = \int_{N_w^-} W(gn)\psi^{-1}(n)dn \tag{6.5}$$

for every $g \in NA_w w N_w^-$.

Remark 6.7. It follows from (6.5) that $j_{\pi,\psi,w}(g)$ satisfies

$$j_{\pi,\psi,w}(n_1 g n_2) = \psi(n_1)\psi(n_2)j_{\pi,\psi,w}(g), \quad \text{for all } n_1 \in N, n_2 \in N_w^-. \tag{6.6}$$

Remark 6.8. If π is supercuspidal then it follows from Remark 4.2 that $\mathcal{W}^0(\pi,\psi) = \mathcal{W}(\pi,\psi)$. Hence $j_{\pi,\psi,w}$ is defined by the equation

$$\int_{N_w^-} W(gn)\psi^{-1}(n)dn = j_{\pi,\psi,w}(g)W(e)$$

for all $W \in \mathcal{W}(G,\psi)$ and $g \in NA_w w N_w^-$.

Lemma 6.9. *$j_{\pi,\psi,w}$ is locally constant on $NA_w w N_w^-$.*

Proof. Using the invariance of $j_{\pi,\psi,w}$ given by (6.6) it is enough to show that $j_{\pi,\psi,w}$ is locally constant on $A_w w$. We write $j_{\pi,\psi,w}$ as an integral of a function $W \in \mathcal{W}^0(\pi,\psi)$ as in (6.5). By Theorem 4.6 $n \mapsto W(awn)$ is compactly supported on N_w^- uniformly for $a \in A^M(w)$ for every positive constant M. Since the sets $A^M(w)$ are open sets of A which cover A the result is clear. $\square$

We finish this section by describing the Bessel functions associated to the contragredient representation. Let π be an irreducible admissible representation of G with a ψ Whittaker functional L as above. Let $\hat{\pi}$ be the representation contragredient to π. Let τ be the involution of G defined by

$$\tau(g) = w_0 {}^t g^{-1} w_0.$$

For each $W \in \mathcal{W}(\pi,\psi)$ we let $W^\tau(g) = W(\tau(g))$. By [**J-PS-S**] the mapping $W \mapsto W^\tau$ is a bijection between $\mathcal{W}(\pi,\psi)$ and $\mathcal{W}(\hat{\pi},\psi^{-1})$.

Lemma 6.10. *The mapping above induces a bijection between $\mathcal{W}^0(\pi,\psi)$ and $\mathcal{W}^0(\hat{\pi},\psi^{-1})$.*

Proof. If $g \in G$ is written in the Iwasawa decomposition in the form $g = nak$ where n is upper triangular a is diagonal and $k \in \mathrm{GL}_3(O)$ then

$$\tau(g) = (w_0 {}^t n^{-1} w_0)(w_0 a^{-1} w_0)(w_0 {}^t k^{-1} w_0)$$

is an Iwasawa decomposition for $\tau(g)$. If $w \in W$ and $g \in BwB$ then $\tau(g) \in Bw_0 w w_0 B$. It is easy to check that $\alpha_{1,2}(a) = \alpha_{2,3}(w_0 a^{-1} w_0)$ and that $\alpha_{1,2} \in S^0(w)$ if and only if $\alpha_{2,3} \in S^0(w_0 w w_0)$. Hence if $W \in \mathcal{W}^0(\pi,\psi)$ and $W^\tau(g) \neq 0$ we get that each $\alpha \in S^0(w_0 w w_0)$ is bounded on $w_0 a^{-1} w_0$ hence each $\alpha \in S^0(w)$ is bounded on a. $\square$

Lemma 6.11. $j_{\hat{\pi},\psi^{-1},w}(g) = j_{\pi,\psi,\tau(w)}(\tau(g)), \quad g \in NA_w wN.$

Proof. Let $W \in \mathcal{W}^0(\pi,\psi)$ such that $W(e) = 1$. (There exists such W by Remark 6.6.) Then by Lemma 6.10, $W^\tau \in \mathcal{W}^0(\hat{\pi},\psi^{-1})$ and $W^\tau(e) = 1$. By (6.5) we have

$$\int_{N_w^-} W^\tau(gn)\psi(n)dn = j_{\hat{\pi},\psi^{-1},w}(g).$$

On the other hand

$$\begin{aligned}\int_{N_w^-} W^\tau(gn)\psi(n)dn &= \int_{N_w^-} W(\tau(gn))\psi(n)dn \\ &= \int_{N_w^-} W(\tau(g)\tau(n))\psi(n)dn \\ &= \int_{N_{w_0ww_0}^-} W(\tau(g)n)\psi^{-1}(n)dn \\ &= j_{\pi,\psi,\tau(w)}(\tau(g)).\end{aligned}$$

□

Corollary 6.12. $j_{\hat{\pi},\psi^{-1}}(g) = j_{\pi,\psi}(g^{-1}), \quad g \in Bw_0B.$

Proof. Recall that $j_{\pi,\psi}(g) = j_{\pi,\psi,w_0}(g)$. Since $\tau(w_0) = w_0$, it follows from Lemma 6.11 that $j_{\hat{\pi},\psi^{-1}}(g) = j_{\pi,\psi}(\tau(g))$. It is clear that both the functions on the right-hand side and of the left-hand side of the equality we are trying to prove are equivariant from the left and right by (N,ψ^{-1}). Hence it is enough to prove the equality for $g = aw_0$. If $g = aw_0$ then $\tau(g) = g^{-1}$ and we are done. □

7. Orbital integrals.

In this section we show that the Bessel functions defined in Section 6 are given locally by orbital integrals. These integrals were studied in [**J-Y1**]. We will do this in two steps. We will show that the Bessel function restricted to a small open set around an arbitrary element $g \in G$ is given by an integral of a Whittaker function which is compactly supported mod N. That is, if we restrict ourselves to this small neighborhood, we can replace a Whittaker function in the representation space with a different Whittaker function (not necessarily in the representation space) which is compactly supported mod N. Then we use the fact that each Whittaker function which is compactly supported mod N comes from an integral of a function in $C_c^\infty(G)$. We will start from the second part.

Let ω be a character of Z and let $\mathcal{W}_\omega(G,\psi) \subseteq \mathcal{W}(G,\psi)$ be the subspace of functions $W \in \mathcal{W}(G,\psi)$ satisfying

$$W(gz) = \omega(z)W(g) \quad g \in G, z \in Z. \tag{7.1}$$

Let $C_c^\infty(G)$ be the space of locally constant functions on G with compact support. For each $f \in C_c^\infty(G)$ we let

$$W_f(g) = W_f^\psi(g) = \int_N f(ng)\psi^{-1}(n)dn.$$

It is clear that $W_f \in \mathcal{W}(G,\psi)$. We also define

$$W_{f,\omega}(g) = \int_N f(nzg)\psi^{-1}(n)\omega^{-1}(z)dndz, \quad f \in C_c^\infty(G). \tag{7.2}$$

It is clear that $W_{f,\omega} \in \mathcal{W}_\omega(G,\psi)$. The image of these maps is well-known. We thank H. Jacquet for providing us with the following proof which we include for the sake of completeness:

Lemma 7.1. *Let $f \in C_c^\infty(G)$. Then W_f is compactly supported* mod N *and the map $f \mapsto W_f$ is a linear map onto the space of compactly supported functions* mod N *in $\mathcal{W}(G,\psi)$.*

Proof. The only nontrivial claim is that the map is onto. Assume $W \in \mathcal{W}(G,\psi)$ is compactly supported mod N. We would like to show that $W = W_f$ for some $f \in C_c^\infty(G)$. Since W is smooth there exist an open compact group U such that $W(gu) = W(u)$ for all $g \in G$ and $u \in U$. Since W is compactly supported mod N, it follows that it is supported on a finite number of double cosets of the form NgU. Hence, it is enough to prove the result for W which is supported on one double coset NgU. For such W we have that $W(ngu) = \psi(n)W(g)$ for all $n \in N$ and $u \in U$ and $W(x) = 0$ if $x \notin NgU$. Assume $W(g) \neq 0$. Then $\psi(n) = 1$ for all $n \in gUg^{-1} \cap N$. Let F be the characteristic function of $gUg^{-1} \cap N$ and let $c = \int_N F(n)dn$. Define $f \in C_c^\infty(G)$ by $f(ngu) = c^{-1}W(g)F(n)$, and $f(x) = 0$ if $x \notin NgU$. It is easy to check that f is well-defined and that $W_f = W$. $\square$

The same proof works for our second projection:

Lemma 7.2. *Let $f \in S(G)$. Then $W_{f,\omega}$ is compactly supported* mod NZ *and the map $f \mapsto W_{f,\omega}$ is a linear map onto the space of compactly supported functions* mod N *in $\mathcal{W}_\omega(G,\psi)$.*

Let $|V|$ be the subspace of $|X|$ given by $|V| = \{|\alpha|_{r_1,r_2,r_3} : r_1+r_2+r_3 = 0\}$. (See (2.7).) Let $Q = \{\beta_1, \beta_2\}$ be a basis for $|V|$. Let $C_1 < C_2$ be positive constants and define

$$A_Q(C_1,C_2) = \{a \in A : \; C_1 < \beta_i(a) < C_2, \; i = 1,2\}.$$

Lemma 7.3. *A function W on G is compactly supported* mod NZ *if and only if there exist constants C_1, C_2 such that W is supported on $NA_Q(C_1, C_2)K$.*

Proof. We can write $A_Q(C_1, C_2) = ZA'$ where

$$A' = \{d(a_1, a_2, 1) \in A_Q(C_1, C_2)\}.$$

Since Q is a basis it is clear that A' is compact. Hence if W is supported on $NA_Q(C_1, C_2)K$ then it is compactly supported mod NZ. Now assume W is compactly supported mod NZ. Then W is supported on a set of the form NZR for some compact set R. Since the sets $NA_Q(C_1, C_2)K$ for different choices of C_1 and C_2 are open sets that cover G we get that the sets of the form $NA_Q(C_1, C_2)K$ cover R. Since R is compact there exist constants C_1', C_2' so that $R \subset NA_Q(C_1', C_2')K$. Hence $NZR \subset NA_Q(C_1', C_2')K$. □

For each $w \in \mathbb{W}$ we define the set $M(w) \subset \Delta^*$ as follows:

$$M(w) = \{\alpha^* | \alpha \in \Delta,\ \alpha \notin S^0(w)\}.$$

Then $M(e) = M(w_{(1,2)}) = M(w_{2,3}) = \emptyset$, $M(w_{1,2,3}) = \{\lambda_2\}$, $M(w_{1,3,2}) = \{\lambda_1\}$ and $M(w_{(1,3)}) = \Delta^*$.

Remark 7.4. It is easily checked that the set $M(w) \cup S^0(w)$ is a basis for $|V|$.

Let E be a positive constant. We let

$$A_{M(w)}(E) = \{a \in A :\ |\lambda|(a) > E \text{ for every } \lambda \in M(w)\}.$$

Theorem 7.5. *Let $W \in \mathcal{W}^0$, $w \in \mathbb{W}$ and $E > 0$. There exist a function $F \in \mathcal{W}^0$ compactly supported* mod NZ *such that*

$$F(n_1 awn_2) = W(n_1 awn_2) \tag{7.3}$$

for all $a \in A_{M(w)}(E)$ and $n_1, n_2 \in N$.

Remark 7.6. Let $C_1 < C_2$ be positive constants and let $A_{C_1,C_2} = A_\Delta(C_1, C_2)$. By Lemma 7.3 we have that F being compactly supported mod NZ is equivalent to F being supported on a set of the form $NA_{C_1,C_2}K$ for some C_1, C_2. Hence we can find $F \in \mathcal{W}^0$ compactly supported mod NZ such that (7.3) holds if and only if we can find constants C_1, C_2 such that the function

$$F(g) = \begin{cases} W(g), & \text{if } g \in NA_{C_1,C_2}K; \\ 0, & \text{otherwise,} \end{cases} \tag{7.4}$$

satisfies (7.3). Hence, we shall use (7.4) to define the desired F. Notice that if we define F by (7.4) then $W(g) = 0 \Rightarrow F(g) = 0$ hence we only need to prove (7.3) for $g = n_1 awn_2$ such that $a \in A_{M(w)}(E)$ and $W(g) \neq 0$.

Proof. We proceed with a case by case analysis as in the proofs of Theorem 3.1, Theorem 4.6 and Theorem 5.7.

Case 1. $w = e$.

In this case $S^0(w) = \Delta$, $M(w) = \emptyset$ and $N_w^- = \{e\}$. Since $W \in \mathcal{W}^0$ it follows that the support of W on B is contained in a set of the form NA_{C_1,C_2}. Define F as in (7.4). Then F satisfies the requirements of the theorem.

Case 2. $w = w_{(1,2)}$.

In this case $S^0(w) = \Delta$ and $M(w) = \emptyset$. Since $W \in \mathcal{W}^0$ it follows that the support of W on BwB is contained in $NA_{C_1,C_2}K$ for some constants C_1, C_2. Hence, if we define F by (7.4) then $F = W$ on BwB. Thus F will satisfy the requirements of the theorem.

Case 3. $w = w_{(2,3)}$.

This case is the same as Case 2.

Case 4. $w = w_{(1,2,3)}$.

In this case $S^0(w) = \{|\alpha|_{2,3} = |\alpha|_{0,1,-1}\}$, $M(w) = \{|\alpha|_{2,-1,-1}\}$ and $N_w^- = N_{2,3}N_{1,3}$.

We consider a special case. Assume that $W(awx_{1,3}(b)) \neq 0$ and $|\alpha|_{2,-1,-1}(a) > E$ (that is, $a \in A_{M(w)}(E)$). If $|b|$ is large then $x_{3,1}(-b^{-1})$ stabilizes W and we have

$$W(awx_{1,3}(b)) = W(awx_{1,3}(b)x_{3,1}(-b^{-1})). \tag{7.5}$$

Now $awx_{1,3}(b)x_{3,1}(-b^{-1}) \in Bw_{(2,3)}B$ and by Case 3 we can choose F as in (7.4) so that $F(g) = W(g)$ for every $g \in Bw_{(2,3)}B$. Since F is smooth F satisfies an equation such as (7.5) for $|b|$ large enough hence we have that there exist a constant D such that

$$F(awx_{1,3}(b)) = W(awx_{1,3}(b)) \tag{7.6}$$

for every $|b| > D$.

Now assume $|b| \leq D$. It follows from the proof of Lemma 4.4 (a) that there exist a constant E_1 such that $|\alpha|_{2,-1,-1}(awx_{1,3}(b)) > E_1$ for every $|b| \leq D$ and every a such that $|\alpha|_{2,-1,-1}(a) > E$. Since we assumed that $W(awx_{1,3}(b)) \neq 0$ and since $|\alpha|_{2,-1,-1}$ is a positive linear combination of $|\alpha|_{1,2}$ and $|\alpha_{2,3}|$ it follows from [**J-PS-S**] (see also Remark 4.2) that there exist a constant E_2 such that $|\alpha|_{2,-1,-1}(awx_{1,3}(b)) < E_2$. Since $W \in \mathcal{W}^0$ there exist constants D_1, D_2 such that $D_1 < |\alpha|_{0,1,-1}(awx_{1,3}(b)) < D_2$. Hence it follows from Lemma 7.3 that $awx_{1,3}(b) \in NA_{C_1,C_2}K$ for some constants C_1, C_2 and if we define F by (7.4) we get that

$$F(awx_{1,3}(b)) = W(awx_{1,3}(b)) \tag{7.7}$$

for every $a \in A$ such that $|\alpha|_{2,-1,-1}(a) > E$ and every $b \in k$ such that $|b| \leq D$. From (7.6), and (7.7) it follows that we can choose the constants C_1 and C_2 to cover both cases and then the F defined by (7.4) will satisfy

$$F(awx_{1,3}(b)) = W(awx_{1,3}(b)) \tag{7.8}$$

for every $a \in A$ such that $|\alpha|_{2,-1,-1}(a) > E$ and every $b \in k$.

For the general case we consider the values of the function

$$W(awx_{1,3}(b_1)x_{2,3}(b_2)),$$

with $|\alpha|_{2,-1,-1}(a) > E$. If $|b_2|$ is very large then we argue as in the special case that there exist constants C_1, C_2 and a function F given by (7.4) such that

$$F(awx_{1,3}(b_1)x_{2,3}(b_2)) = W(awx_{1,3}(b_1)x_{2,3}(b_2)) \tag{7.9}$$

for every for every $a \in A$ such that $|\alpha|_{2,-1,-1}(a) > E$ and every $b_2 \in k$ such that $|b_2|$ is larger than some constant D.

Assume $|b_2| \leq D$. There exist functions $W_1, \dots, W_l \in \mathcal{W}^0$ such that

$$\rho(x_{2,3}(b_2))W = c_1(b)W_1 + \cdots + c_l(b)W_l$$

for all $|b_2| \leq D$ and c_i are locally constant functions from $\{x : |x| \leq C\}$ to $\mathbb{C}$. By our special case, there exist functions $F_1, \dots, F_l \in \mathcal{W}^0$ compactly supported mod NZ such that

$$F_i(awx_{1,3}(b_1)) = W_i(awx_{1,3}(b_1))$$

for every $a \in A$ such that $|\alpha|_{2,-1,-1}(a) > E$. Let $y \in k$ with $|y| \leq D$ and define

$$F_y = c_1(y)\rho(x_{2,3}(-y))F_1 + \cdots + c_l(y)\rho(x_{2,3}(-y))F_l.$$

Then we have

$$F_y(awx_{1,3}(b_1)x_{2,3}(y)) = W(awx_{1,3}(b_1)x_{2,3}(y)) \tag{7.10}$$

for every $a \in A$ such that $|\alpha|_{2,-1,-1}(a) > E$. Since every function in sight is locally constant the equality in (7.10) will remain true if we fix F_y but change $x_{2,3}(y)$ to $x_{2,3}(y')$ for some y' in a small neighborhood of y. Since F_y is compactly supported mod NZ it follows from Remark 7.6 that we can choose F as in (7.4) so that (7.9) will hold for every a as above and every $|b_2| \leq D$ in a small neighborhood of y. Since we need only a finite number of such y to cover the set $\{b_2 : |b_2| \leq D\}$ we can change the constants C_1 and C_2 in the definition of F in (7.4) to get the conclusion of the theorem.

Case 5. $w = w_{(1,3,2)}$.

This case is similar to Case 4 and is omitted.

Case 6. $w = w_0 = w_{(1,3)}$.

In this case $S^0(w) = \emptyset$ hence $M(w) = \Delta^* = \{\lambda_1, \lambda_2\}$. Also $N_w^- = N$. We first consider two special cases. Since the arguments are the same as in Case 4 and as in our general case below we omit them.

In the first special case we prove that there exist constants C_1, C_2 depending on E such that if F is given by (7.4) then

$$F(awx_{1,3}(b)) = W(awx_{1,3}(b))$$

for every $a \in A$ such that $\lambda_1(a) > E$ and $\lambda_2(a) > E$. The proof of this case is the same as the proof of the special case of Case 4.

In the second special case we prove that there exist constants C_1, C_2 depending on E such that if F is given by (7.4) then

$$F(awx_{1,3}(b_1)x_{2,3}(b_2)) = W(awx_{1,3}(b_1)x_{2,3}(b_2))$$

for every $a \in A$ such that $\lambda_1(a) > E$ and $\lambda_2(a) > E$. The proof of this case uses the first special case and is similar to our general case below.

We consider the function $W(awx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3))$. If $|b_3|$ is very large then

$$W(awx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3)) = W(awx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3)x_{2,1}(-b_3^{-1})).$$

Let $g = awx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3)x_{2,1}(-b_3^{-1})$. Then we can write $g = n_1\widetilde{a}w_{(1,2,3)}n_2$ with $n_1 \in N$, $\widetilde{a} \in A$ and $n_2 \in N^-_{w_{(1,2,3)}}$. It is also easy to see that $\widetilde{a} = ah_{3,2}(b_2)$. Since $\lambda_1(ah_{3,2}(b_2)) = \lambda_1(a)$ it follows from Case 4 that there exist constants C_1, C_2 depending on E such that if F is defined by (7.4) then

$$F(n_1ah_{3,2}(b_2)w_{(1,2,3)}n_2) = W(n_1ah_{3,2}(b_2)w_{(1,2,3)}n_2)$$

for every $|a|$ such that $\lambda_1(a) > E$. Since F is smooth it follows that there exist a constant $|D|$ such that

$$F(awx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3)) = W(awx_{1,3}(b_1)x_{2,3}(b_2)x_{1,2}(b_3)) \tag{7.11}$$

for every $a \in A$ and $b_3 \in k$ such that $\lambda_1(a) > E$ and $|b_3| > D$. For the case where $|b_3| \le D$ we argue as in Case 4 using our special case above. $\square$

For $W \in \mathcal{W}^0$ and $g \in Bw_0B$ we let

$$J(W,g) = J_{g,w_0}(W) = \int_N W(gn)\psi^{-1}(n)dn.$$

Corollary 7.7. *Let $W \in \mathcal{W}^0$ and $x \in G$. Let U_x be a compact neighborhood of x in G. Then there exists a function $F \in \mathcal{W}$ compactly supported* mod *NZ such that*

$$J(F,g) = J(W,g)$$

for every $g \in U_x \cap Bw_0B$.

Proof. Let $\Delta^* = \{\lambda_1, \lambda_2\}$ as above. Since U_x is compact it follows that λ_1 and λ_2 are both bounded on U_x. Hence there exist a constant E such that $\lambda_1(g) > E$ and $\lambda_2(g) > E$ for every $g \in U_x$. Let $g \in U_x \bigcap Bw_0B$. Then $g = n_1aw_0n_2$ for some $n_1, n_2 \in N$ and $a \in A$. By Theorem 3.1 we have

$$\lambda_i(g) = \lambda_i(n_1aw_0n_2) = \lambda_i(a)\lambda(w_0n_2) \le \lambda_i(a)$$

for $i = 1, 2$. Since $\lambda_i(g) > E$ it follows that $\lambda_i(a) > E$ for $i = 1, 2$. By Theorem 7.5 we can find F in $\mathcal{W}$ compactly supported mod NZ such that

$$F(aw_0n) = W(aw_0n)$$

for every $n \in N$ and $a \in A$ such that $\lambda_i(a) > E$, $i = 1,2$. In particular we have $F(gn) = W(gn)$ for every $g \in U_x \cap Bw_0B$ and $n \in N$. Hence $J(F,g) = J(W,g)$ for such g which is what we wanted to prove. $\square$

Let $\phi \in C_c^\infty(G)$ and ω a quasi-character of Z. Let $g \in Bw_0B$. We define

$$J_{\psi,\omega}(\phi,g) = \int_N \int_N \int_Z f(n_1zgn_2)\psi^{-1}(n_1)\psi^{-1}(n_2)\omega^{-1}(z)dn_1dn_2dz.$$

It follows from [**J-Y1**] that this integral converges absolutely. It is easy to see that

$$J_{\psi,\omega}(\phi,g) = \int_N W_{\phi,\omega}(gn)\psi^{-1}(n)dn = J(W_{\phi,\omega},g)$$

where $W_{\phi,\omega}$ is defined by (7.2).

Corollary 7.8. *Let π be an irreducible admissible representation of G with central character ω_π. Let $x \in G$ and let U_x be a compact neighborhood of x in G. Then there exist a function $\phi \in C_c^\infty(G)$ such that*

$$J_{\psi,\omega_\pi}(\phi,g) = j_{\pi,\psi}(g) \tag{7.12}$$

for every $g \in U_x \cap Bw_0B$.

Proof. Recall that $j_{\pi,\psi}(g) = j_{\pi,\psi,w_0}(g)$ is the Bessel function associated to the longest Weyl element w_0. By (6.5) there exist $W \in \mathcal{W}^0(\pi,\psi)$ such that $j_{\pi,\psi}(g) = J(W,g)$ for every $g \in Bw_0B$. Since the central character of π is ω_π it follows that $W \in \mathcal{W}^0_{\omega_\pi}(G,\psi) = \mathcal{W}_{\omega_\pi}(G,\psi) \cap \mathcal{W}^0(G,\psi)$. By Corollary 7.7 there exist $F \in \mathcal{W}(G,\psi)$ which is compactly supported mod NZ such that $J(F,g) = J(W,g)$ for every $g \in U_x \cap Bw_0B$. It is easy to see from the proof of Theorem 7.5 that if $W \in \mathcal{W}^0_\omega(G,\psi)$ then F can be chosen in $\mathcal{W}_\omega(G,\psi)$ (see (7.4)). By Lemma 7.2 there exist $\phi \in C_c^\infty(G)$ such that $W_{\phi,\omega} = F$. Hence

$$J_{\psi,\omega}(\phi,g) = J(W_{\phi,\omega},g) = J(F,g) = J(W,g) = j_{\pi,\psi}(g)$$

for every $g \in U_x \cap Bw_0B$. $\square$

Remark 7.9. If π is supercuspidal then there exis $\phi \in C_c^\infty(G)$ such that (7.12) holds for every $g \in Bw_0B$. That is, in this case the Bessel function is globally given by an orbital integral. This follows from Remark 4.2, from Lemma 7.2 and (6.5).

Acknowledgments. I thank J. Cogdell, H. Jacquet, I. Piatetski-Shapiro, and S. Rallis for sharing their insight with me and their constant encouragement.

References

[Av] Z. Averbuch, *Bessel functions of algebraic groups' representations over a local field*, Tel Aviv University, 1992.

[Ba] E.M. Baruch, *On Bessel distributions for quasi-split groups*, Trans. Amer. Math. Soc., **353**(**7**) (2001), 2601-2614, MR 2002b:22029, Zbl 0976.22015.

[Ba2] ______, *Bessel distributions for $GL(3)$ over the p-adics*, Pacific J. Math., to appear.

[Bo] A. Borel, *Introduction aux Groupes Arithmétiques*, Hermann, Paris, 1969, MR 39 #5577, Zbl 0186.33202.

[Bo-HC] A. Borel and Harish-Chandra, *Arithmetic subgroups of algebraic groups*, Ann. of Math. (2), **75** (1962), 485-535, MR 26 #5081, Zbl 0107.14804.

[C-PS1] J.W. Cogdell and I. Piatetski-Shapiro, *The Arithmetic and Spectral Analysis of Poincaré Series*, Perspectives in Mathematics, **13**, Academic Press Inc., Boston, MA, 1990, MR 91h:11042, Zbl 0714.11032.

[G-PS] S. Gelbart and I. Piatetski-Shapiro, *Distinguished representations and modular forms of half-integral weight*, Invent. Math., **59**(**2**) (1980), 145-188, MR 82b:10035, Zbl 0426.10027.

[HC] Harish-Chandra, *Admissible Invariant Distributions on Reductive p-adic Groups*, University Lecture Series, **16**, American Mathematical Society, Providence, RI, 1999, MR 2001b:22015, Zbl 0433.22012

[H] R. Howe, *Classification of irreducible representations of* $\mathrm{GL}_2(F)$*, F a local field*, preprint, 1978.

[J-PS-S] H. Jacquet, I. Piatetski-Shapiro and J. Shalika, *Rankin-Selberg convolutions*, Amer. J. Math., **105**(**2**) (1983), 367-464, MR 85g:11044, Zbl 0525.22018.

[J-Y1] H. Jacquet and Y. Ye, *Distinguished representations and quadratic base change for* GL(3), Trans. Amer. Math. Soc., **348**(**3**) (1996), 913-939, MR 96h:11041, Zbl 0861.11033.

[PS] I. Piatetski-Shapiro, *Complex Representations of* $\mathrm{GL}(2,K)$ *for Finite Fields K*, Contemporary Math., **16**, AMS, 1983, MR 84m:20046, Zbl 0513.20026.

[Rod] F. Rodier, *Modéle de Whittaker et caract'eres de repr'esentations*, Lecture Notes in Mathematics, **466**, Springer Verlag, 1974, 151-171, MR 52 #14165, Zbl 0339.22014.

[S] D. Soudry, *The L and γ factors for generic representations of $GSp(4,k) \times$* GL(2, k) *over a local non-Archimedean field k*, Duke Math. J., **51**(**2**) (1984), 355-394, MR 86f:22022, Zbl 0557.12012.

Received February 4, 2002. Partially supported by NSF grant DMS-0070762.

Department of Mathematics
Technion
Haifa 32000
Israel
E-mail address: embaruch@math.technion.ac.il

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

A PROPERTY OF FREE ENTROPY

S.T. Belinschi and H. Bercovici

We show that the restriction on the uniform norms of approximating matricial microstates can be removed when defining free entropy.

1. Introduction.

Denote by $\mathfrak{M}_k$ the algebra of complex $k \times k$ matrices, and by τ_k the normalized trace on $\mathfrak{M}_k$, i.e., $\tau_k(A) = \frac{1}{k}\mathrm{Tr}(A)$ for $A \in \mathfrak{M}_k$. Consider for each k a standard Gaussian Hermitian random matrix X_k. Thus, if E denotes expected value, $E\tau_k(X_k) = 0$ and $E\tau_k(X_k^2) = 1$. It was shown by E. Wigner [**9**] that, as $k \to \infty$, X_k tends in distribution to a semicircular law, i.e., the limits

$$\mu_p = \lim_{k\to\infty} E\tau_k(X_k^p)$$

exist, and they can be calculated as

$$\mu_p = \frac{1}{2\pi}\int_{-2}^{2} t^p\sqrt{4-t^2}\,dt$$

for $p = 1, 2, \ldots$. If we have several independent standard Gaussian Hermitian random matrices $(X_k(i))_{i=1}^n$, D. Voiculescu [**4**] proved that, as $k \to \infty$, these sets of variables converge in distribution to a free semicircular family. Briefly, this means that given indices $i_j \in \{1, 2, \ldots, n\}$ such that $i_j \neq i_{j+1}$ for $j = 1, 2, \ldots, m-1$, and given positive integers $p_1, p_2, \ldots, p_m$, the limit

$$\lim_{k\to\infty} E\tau_k(X_k(i_1)^{p_1} X_k(i_2)^{p_2} \ldots X_k(i_m)^{p_m})$$

exists, and

$$\lim_{k\to\infty} E\tau_k[(X_k(i_1)^{p_1} - \mu_{p_1})(X_k(i_2)^{p_2} - \mu_{p_2}) \ldots (X_k(i_m)^{p_m} - \mu_{p_m})] = 0.$$

It is natural to look for large deviation principles associated with these limit laws. For this purpose (and also with motivation from information theory and statistical physics) Voiculescu introduced in [**6**] (cf. also [**5**]) the notion of free entropy. The original definition of free entropy, which will be reviewed below, involves a bound $R > 0$ on the operator norm of approximating matricial microstates, and this may perhaps obscure its significance for large deviations. It is our purpose here to show that this bound can be removed — roughly speaking, one can set $R = \infty$ in the definition of free entropy. This result applies to other notions of free entropy which appeared subsequently (see for instance [**7**] for free entropy in the presence of additional variables, [**8**] for free entropy using an ultrafilter, and [**3**] for free entropy of a nonselfadjoint variable). We will only provide the proof for the original quantity χ defined in [**6**], but it should be obvious how the argument applies in the other situations.

It should be noted that a large deviation theorem for Wigner's result has been proved by G. Ben Arous and A. Guionnet [**1**], where the natural topology of weak convergence of probability measures on the real line is used. The rate function is closely related with free entropy. For several variables, a thorough study of large deviations was undertaken by T. Cabanal Duvillard and A. Guionnet [**2**]. The rate function they determine is related with another version of free entropy (microstate free).

2. The main result.

For the remainder of this note we fix a positive integer n. We will denote by I the collection of all multiindices $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$ with $m \geq 1$ and $\alpha_j \in \{1, 2, \dots, n\}$ for all $j = 1, 2, \dots, m$. In other words, $I = \bigcup_{m=1}^{\infty} \{1, 2, \dots, n\}^m$. A multiindex of the form $(\alpha, \alpha, \dots, \alpha)$ will also be denoted α^m. We consider the space $\mathbb{S}$ consisting of all families $(\mu(\alpha))_{\alpha \in I}$ of complex numbers indexed by I. The space $\mathbb{S}$ will be endowed with the topology of componentwise convergence.

Consider now a tracial W^*-probability space $(\mathfrak{A}, \tau)$. That is, $\mathfrak{A}$ is a von Neumann algebra, and τ is a normal trace state on $\mathfrak{A}$. We will write $\mathfrak{A}^{\mathrm{sa}}$ for the space of selfadjoint elements of $\mathfrak{A}$. Given an n-tuple $X = (X_1, X_2, \dots, X_n) \in (\mathfrak{A}^{\mathrm{sa}})^n$, its distribution $\mu_X \in \mathbb{S}$ is defined by

$$\mu_X(\alpha) = \tau(X_\alpha),$$

where $X_\alpha = X_{\alpha_1} X_{\alpha_2} \dots X_{\alpha_m}$ for $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m) \in I$. This notation applies in particular to n-tuples of selfadjoint matrices in $\mathfrak{M}_k$. Voiculescu's entropy measures the extent to which the distribution of X can be approximated by distributions of the form μ_A with $A \in (\mathfrak{M}_k^{\mathrm{sa}})^n$. Note first that $\mathfrak{M}_k^{\mathrm{sa}}$ is a real Hilbert space with the Hilbert-Schmidt norm $\|A\|_2 = \mathrm{Tr}(A^2)$, and λ_k will denote the corresponding Lebesgue measure (i.e., a cube whose sides form an orthonormal basis has measure equal to one). On the space $(\mathfrak{M}_k^{\mathrm{sa}})^n$ we have the product measure $\lambda_k^{\otimes n}$.

Given $X \in (\mathfrak{A}^{\mathrm{sa}})^n$, and a neighborhood U of μ_X in $\mathbb{S}$, we set

$$\Gamma(X; k, U) = \{A \in (\mathfrak{M}_k^{\mathrm{sa}})^n : \mu_A \in U\}.$$

Given in addition a positive number R,

$$\Gamma_R(X; k, U) = \{A \in \Gamma(X; k, U) : \|A_j\| < R \text{ for all } j\}.$$

We can then define the quantities

$$\chi_R(X;U) = \liminf_{k\to\infty} \left[\frac{1}{k^2} \log \lambda_k^{\otimes n}(\Gamma_R(X;k,U)) + \frac{n}{2}\log k \right],$$

and

$$\chi_R(X) = \inf_U \chi_R(X;U),$$

where U runs over a neighborhood base of μ_X in $\mathbb{S}$. Finally, the free entropy is defined as

$$\chi(X) = \sup_{R>0} \chi_R(X).$$

We also set

$$\chi_\infty(X;U) = \liminf_{k\to\infty} \left[\frac{1}{k^2} \log \lambda_k^{\otimes n}(\Gamma(X;k,U)) + \frac{n}{2}\log k \right],$$

and $\chi_\infty(X) = \inf_U \chi_\infty(X;U)$. This quantity was introduced in the concluding remarks of [**6**], where other possible definitions of free entropy are discussed briefly. The inequalities

$$\chi_R(X) \le \chi(X) \le \chi_\infty(X)$$

are obvious for $R > 0$, and Proposition 2.4 of [**6**] states that $\chi_R(X) = \chi(X)$ if R is sufficiently large; $R > \max_j \|X_j\|$ will suffice. Our main result is as follows:

Proposition 2.1. *For every $X \in (\mathfrak{A}^{\mathrm{sa}})^n$ we have $\chi(X) = \chi_\infty(X)$.*

The proof of this result is a refinement of the proof of Proposition 2.4 in [**6**]. We begin by considering the diffeomorphism f of the real line onto $(-2,2)$ defined by $f(t) = t$ for $t \in [-1,1]$, $f(t) = 2 - \frac{1}{t}$ for $t > 1$, and $f(t) = -2 - \frac{1}{t}$ for $t < -1$. Observe that f' does not have any local minimum, and therefore

$$\frac{f(s) - f(t)}{s-t} \ge \min\{f'(s), f'(t)\} \ge f'(s)f'(t)$$

for all s and t. The function $F_n : (\mathfrak{M}_k^{\mathrm{sa}})^n \to (\mathfrak{M}_k^{\mathrm{sa}})^n$ defined by

$$F_n(A_1, A_2, \ldots, A_n) = (f(A_1), f(A_2), \ldots, f(A_n))$$

is also differentiable, and we need to estimate the Jacobian determinant $(JF_n)(A)$. Since

$$(JF_n)(A) = (JF_1)(A_1)(JF_1)(A_2)\ldots(JF_1)(A_n),$$

it suffices to do this in one variable. As pointed out in [**6**], if A is a $k \times k$ matrix with eigenvalues $\mu_1, \mu_2, \ldots, \mu_k$, we have

$$(JF_1)(A) = \left(\prod_{i\ne j} \frac{f(\mu_i) - f(\mu_j)}{\mu_i - \mu_j} \right) \cdot \prod_{i=1}^k f'(\mu_i).$$

By the estimate for difference quotients shown above,

$$
\begin{aligned}
(JF_1)(A) &\geq \left(\prod_{i \neq j} f'(\mu_i) f'(\mu_j) \right) \cdot \prod_{i=1}^{k} f'(\mu_i) \\
&= \prod_{i=1}^{k} f'(\mu_i)^{2k-1} = \prod_{|\mu_i|>1} \mu_i^{-2(2k-1)}.
\end{aligned}
$$

Denoting $\log^+(t) = \max\{\log t, 0\}$, we obtain

$$
\begin{aligned}
\log(JF_1)(A) &\geq -2(2k-1) \sum_{i=1}^{k} \log^+ \mu_i \\
&= -2k(2k-1) \frac{1}{k} \sum_{i=1}^{k} \log^+ \mu_i \\
&= -2k(2k-1) \tau_k(\log^+ |A|).
\end{aligned}
$$

We have therefore proved the following estimate:

Lemma 2.2. *Given $A = (A_1, A_2, \ldots, A_n) \in (\mathfrak{M}_k^{\mathrm{sa}})^n$, we have*

$$
(JF_n)(A) \geq \exp \left[-2k(2k-1) \sum_{j=1}^{n} \tau_k(\log^+ |A_j|) \right].
$$

Note for further use that, for a selfadjoint $k \times k$ matrix A, $\tau_k(\log^+ |A|)$ can be estimated in terms of the moments $\tau_k(A^{2p})$, $p \geq 1$. In fact, $\log^+ t = \frac{1}{2p} \log^+ t^{2p} \leq \frac{1}{2p} t^{2p}$, and therefore

$$
\tau_k(\log^+ |A|) \leq \frac{1}{2p} \tau_k(A^{2p}).
$$

We need one more ingredient.

Lemma 2.3. *Let $X \in (\mathfrak{A}^{\mathrm{sa}})^n$ satisfy $\max_j \|X_j\| < 1$, and let U be a neighborhood of μ_X in $\mathbb{S}$. There exists a neighborhood V of μ_X in $\mathbb{S}$ such that*

$$
F_n(\Gamma(X; k, V)) \subset \Gamma_2(X; k, U) \text{ for all } k.
$$

Proof. Clearly it suffices to prove the lemma for neighborhoods of the form

$$
U = \{\mu \in \mathbb{S} : |\mu(\alpha) - \tau(X_\alpha)| < \varepsilon\},
$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m) \in I$ and $\varepsilon > 0$ are fixed. Using the Hölder inequality

$$
\|A_\alpha\|_1 \leq \|A_{\alpha_1}\|_m \|A_{\alpha_2}\|_m \ldots \|A_{\alpha_m}\|_m,
$$

we see that it is sufficient to choose V so that, for all $A \in \Gamma(X; k, V)$, we have $|\tau_k(A_\alpha) - \tau(X_\alpha)| < \varepsilon/2$, $\|A_j\|_m \leq 1$, and $\|A_j - f(A_j)\|_m < \varepsilon/2m$ for

$j = 1, 2, \ldots, n$. Choose a number $r < 1$ so that $r > \|X_j\|$ for all j, and choose an even integer $q > m$ such that $r^{q/m} < \varepsilon/2m$. Define next

$$V = \left\{\mu \in \mathbb{S} : |\mu(\alpha) - \tau(X_\alpha)| < \frac{\varepsilon}{2} \text{ and } |\mu(j^q)| < r^q \text{ for } j = 1, 2, \ldots, n\right\};$$

recall that j^q denotes the q-multiindex with all entries equal to j. Consider now $A \in \Gamma(X; k, V)$, and note that the inequalities $|\tau_k(A_\alpha) - \tau(X_\alpha)| < \varepsilon/2$ are obviously satisfied. Also,

$$\|A_j\|_m \leq \|A_j\|_q = \tau(A_j^q)^{1/q} \leq r < 1.$$

Finally, if $\mu_1, \mu_2, \ldots, \mu_k$ are the eigenvalues of A_j,

$$\begin{aligned}\|A_j - f(A_j)\|_m &\leq \left(\frac{1}{k}\sum_{|\mu_j|\geq 1} |\mu_j|^m\right)^{1/m} \leq \left(\frac{1}{k}\sum_{|\mu_j|\geq 1} |\mu_j|^q\right)^{1/m} \\ &\leq (\tau_k(A_j^q))^{1/m} \leq r^{q/m},\end{aligned}$$

and this quantity is less than $\varepsilon/2m$. □

Proof of Proposition 2.1. It suffices to prove the proposition in case $\|X_j\| < 1$ for all j. From the results of [**6**] we know that $\chi_2(X) = \chi(X)$, and clearly $\chi_2(X) \leq \chi_\infty(X)$. To prove the opposite inequality $\chi_2(X) \geq \chi_\infty(X)$, let U be a neighborhood of μ_X in $\mathbb{S}$, and let V be the neighborhood of μ_X furnished by Lemma 2.3, i.e., $F_n(\Gamma(X; k, V)) \subset \Gamma_2(X; k, U)$ for all $k \geq 1$. Given a positive integer p, we may also assume that $\tau_k(A_j^{2p}) \leq 1$ whenever $A = (A_1, A_2, \ldots, A_n) \in \Gamma(X; k, V)$. It follows then from Lemma 2.2 (and the remark following its statement) that

$$(JF_n)(A) \geq \exp\left[-2k(2k-1)\frac{n}{2p}\right]$$

for all $A \in \Gamma(X; k, V)$. Since the function F_n is one-to-one, we deduce that

$$\begin{aligned}\lambda_k^{\otimes n}(\Gamma_2(X; k, U)) &\geq \lambda_k^{\otimes n}(F_n(\Gamma(X; k, V))) \\ &\geq \exp\left[-2k(2k-1)\frac{n}{2p}\right] \lambda_k^{\otimes n}(\Gamma(X; k, V)).\end{aligned}$$

Therefore

$$\begin{aligned}&\frac{1}{k^2}\log \lambda_k^{\otimes n}(\Gamma_2(X; k, U)) + \frac{n}{2}\log k \\ &\geq \frac{1}{k^2}\log \lambda_k^{\otimes n}(\Gamma(X; k, V)) + \frac{n}{2}\log k - \left(2 - \frac{1}{k}\right)\frac{n}{p},\end{aligned}$$

and as $k \to \infty$ this yields

$$\chi_2(X; U) \geq \chi_\infty(X; V) - \frac{2n}{p}.$$

Since p is arbitrary, we deduce that $\chi_2(X;U) \geq \chi_\infty(X;V) \geq \chi_\infty(X)$, and the proof is concluded by taking the infimum over U. $\square$

We remark that a suitable modification of the above proof yields directly that $\chi_\infty(X) = \chi_R(X)$ if $\|X_j\| < R$. One needs an appropriate version of the function f, and that is easily constructed.

References

[1] G. Ben Arous and A. Guionnet, *Large deviations for Wigner's law and Voiculescu's noncommutative entropy*, Probab. Theory Related Fields, **108** (1997), 517-542, MR 98i:15026, Zbl 0954.60029.

[2] T. Cabanal Duvillard and A. Guionnet, *Large deviations upper bounds for the laws of matrix-valued processes and non-commutative entropies*, Ann. Probab., **29** (2001), 1205-1261, MR 2003a:60040.

[3] P. Śniady, *Inequality for Voiculescu's free entropy in terms of Brown measure*, Int. Math. Res. Notices, (2003), 51-64, CMP 1 935 566.

[4] D. Voiculescu, *Limit laws for random matrices and free products*, Invent. Math., **104** (1991), 201-220, MR 92d:46163, Zbl 0736.60007.

[5] ———, *The analogues of entropy and Fisher's information measure in free probability*, I, Comm. Math. Phys., **155** (1993), 71-92, MR 94k:46137, Zbl 0781.60006.

[6] ———, *The analogues of entropy and Fisher's information measure in free probability*, II, Invent. Math., **1118** (1994), 411-440, MR 96a:46117, Zbl 0820.60001.

[7] ———, *The analogues of entropy and Fisher's information measure in free probability.* III. *The absence of Cartan algebras*, Geom. Funct. Anal., **6** (1996), 172-199, MR 96m:46119, Zbl 0856.60012.

[8] ———, *A strengthened asymptotic freeness result for random matrices with applications to free entropy*, Internat. Math. Res. Notices, (1998), 41-63, MR 2000d:46080, Zbl 0895.60004.

[9] E. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. Math., **67** (1958), 325-327, MR 20 #2029, Zbl 0085.13203.

Received May 6, 2002 and revised June 20, 2002. The second author was supported in part by a grant from the National Science Foundation.

Institute of Mathematics
Romanian Academy
P. O. Box 1-764
Bucharest RO-70700
Romania
E-mail address: sbelinsc@indiana.edu

Mathematics Department
Indiana University
Bloomington, IN 47405
E-mail address: bercovic@indiana.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

COTILTING MODULES OVER TAME HEREDITARY ALGEBRAS

Aslak Bakke Buan and Henning Krause

Dedicated to Raymundo Bautista and Roberto Martínez on the occasion of their sixtieth birthday.

For a left noetherian ring Λ, we establish a bijective correspondence between equivalence classes of cotilting Λ-modules which are not necessarily finitely generated, and torsion pairs $(\mathcal{T}, \mathcal{F})$ for the category of finitely generated Λ-modules with $\Lambda \in \mathcal{F}$. In the second part of this paper, we give a complete classification of all cotilting modules over a tame hereditary artin algebra.

Introduction.

The concept of a tilting module was introduced by Brenner and Butler for the category of finitely generated modules over a finite dimensional algebra [**5**]. More recently, various authors studied tilting and cotilting modules in the category of all modules over arbitrary associative rings. For example, Göbel and Trlifaj classified the cotilting modules over the ring $\mathbb{Z}$ of integers [**14**].

In this paper we concentrate on the representation theory of a finite dimensional algebra Λ. Our aim is to show that *all* cotilting modules are relevant when one studies the category $\operatorname{mod}\Lambda$ of finitely generated modules. More specifically, we establish a correspondence between cotilting Λ-modules which are not necessarily finitely generated, and torsion pairs for $\operatorname{mod}\Lambda$.

In the second part of this paper we give a complete classification of all cotilting modules over a finite dimensional tame hereditary algebra. Note that new cotilting modules which are not equivalent to finitely generated ones arise only if Λ is of infinite representation type. The tame hereditary algebras are of infinite type and their modules are fairly well understood. In fact, the category of modules over a tame hereditary algebra shares a lot of properties with the category of modules over a Dedekind domain. It is therefore tempting to choose the tame hereditary algebras to give for the first time a complete classification of all cotilting modules for a finite dimensional algebra of infinite representation type.

Let us recall from [**7**] the definition of a cotilting module of injective dimension at most 1 for an associative ring Λ. We denote by $\operatorname{Mod}\Lambda$ the category of all left Λ-modules. For a module T, let $\operatorname{Prod} T$ be the category of all direct summands in any product of copies of T. A Λ-module T is called *cotilting module* if the following[1] hold:

(C1) The injective dimension of T is at most 1.
(C2) $\operatorname{Ext}(T^\alpha, T) = 0$ for every cardinal α.
(C3) There exists an injective cogenerator Q and an exact sequence $0 \to T'' \to T' \to Q \to 0$ with T', T'' in $\operatorname{Prod} T$.
(C4) T is pure-injective.

By definition, two cotilting modules T and T' are equivalent if $\operatorname{Prod} T = \operatorname{Prod} T'$.

We have the following general result about cotilting modules:

Theorem A. *Let Λ be a left noetherian ring. There exists a bijection between:*

- *Torsion pairs $(\mathcal{T}, \mathcal{F})$ for $\operatorname{mod}\Lambda$ such that $\Lambda \in \mathcal{F}$, and*
- *equivalence classes of cotilting modules.*

A torsion pair $(\mathcal{T}, \mathcal{F})$ corresponding to a cotilting module T satisfies $\mathcal{F} = \{X \in \operatorname{mod}\Lambda \mid \operatorname{Ext}(X, T) = 0\}$.

From now on assume that Λ is a tame hereditary finite dimensional algebra [**18**]. In order to formulate the classification of all cotilting modules we need to recall a few well-known facts. We denote by $\mathcal{R}$ the category of finitely generated regular modules. This is an abelian category and we write $\mathbb{P}$ for the set of isoclasses of simple objects in $\mathcal{R}$. For each $S \in \mathbb{P}$, let $[S]$ denote the equivalence class of S with respect to the smallest equivalence relation on $\mathbb{P}$ satisfying $S \sim S'$ if $\operatorname{Ext}(S, S') \neq 0$. For each $S \in \mathbb{P}$ and $n \in \mathbb{N}$, let S_n be the unique indecomposable object in $\mathcal{R}$ of length n satisfying $\operatorname{Hom}(S, S_n) \neq 0$, and S_{-n} denotes the unique indecomposable object of length n satisfying $\operatorname{Hom}(S_{-n}, S) \neq 0$. There are chains of monomorphisms $S = S_1 \to S_2 \to \dots$ and chains of epimorphisms $\cdots \to S_{-2} \to S_{-1} = S$ for each $S \in \mathbb{P}$. The corresponding *Prüfer module* is the colimit $S_\infty = \varinjlim S_n$ whereas the *adic* module is $S_{-\infty} = \varprojlim S_{-n}$ (which is often denoted by $\widehat{S}$). Moreover, there is a unique *generic* module G, that is, G is indecomposable of infinite length and has finite length over $\operatorname{End}(G)$.

For a module M, we denote by $\operatorname{indec} M$ the set of isoclasses of indecomposable direct summands of M. If M is pure-injective, then there is a unique family $(M_i)_{i \in I}$ of modules $M_i \in \operatorname{indec} M$ such that M is the pure-injective envelope of $\coprod_i M_i$. The well-known classification of the indecomposable pure-injectives over a tame hereditary algebra allows us to classify all cotilting modules.

[1]Recently, Silvana Bazzoni has shown that (C1)-(C3) implies (C4).

Theorem B. *Let Λ be a tame hereditary algebra and let T be a pure-injective Λ-module.*

(1) *Suppose all indecomposable direct summands are finitely generated. Then T is a cotilting module if and only if the number of non-isomorphic indecomposable direct summands of T equals the number of simple Λ-modules and $\operatorname{Ext}(T', T'') = 0$ for all $T', T'' \in \operatorname{indec} T$.*
(2) *Suppose there is an indecomposable direct summand which is not finitely generated. Then T is a cotilting module if and only if the following hold:*
   - *Each indecomposable direct summand of T is either generic or of the form S_n for some $S \in \mathbb{P}$ and some $n \in \mathbb{N} \cup \{-\infty, \infty\}$.*
   - *For each $S \in \mathbb{P}$, let I_S be the set of non-isomorphic indecomposable direct summands of T which are of the form S'_n for some $n \in \mathbb{N} \cup \{-\infty, \infty\}$ and some $S' \in [S]$. Then $\operatorname{card} I_S = \operatorname{card}[S]$ and $\operatorname{Ext}(T', T'') = 0$ for all $T', T'' \in I_S$.*
(3) *Two cotilting modules T_1 and T_2 are equivalent if and only if $\operatorname{indec}(T_1 \amalg G) = \operatorname{indec}(T_2 \amalg G)$.*

The paper is organized as follows: In Section 1 we establish the bijective correspondence between cotilting modules and torsion pairs. In addition, we include a generalization to locally noetherian Grothendieck categories. From Section 2 onwards, we restrict ourselves to tame hereditary algebras. First, we recall briefly the classification of the pure-injective modules. Then we describe the selforthogonal pure-injective modules. A cotilting module is a selforthogonal module which is maximal with respect to this property. This observation is the basis of our classification of all cotilting modules which is completed in Section 3. Each cotilting module which is not equivalent to a finitely generated one is the pure-injective envelope of a direct sum $\coprod_{\mathcal{T}} T_{\mathcal{T}}$ where $\mathcal{T}$ runs through the set of tubes in the category of regular modules. Each summand $T_{\mathcal{T}}$ can be viewed as a cotilting object in some appropriate Grothendieck category, and a detailed description is given in the final Section 4.

1. A correspondence for cotilting modules.

Let Λ be an associative k-algebra over some commutative ring k. Throughout we assume that Λ is left noetherian. Let $\operatorname{Mod}\Lambda$ be the category of (left) Λ-modules, and let $\operatorname{mod}\Lambda$ denote the full subcategory which is formed by all finitely generated modules. To simplify our notation, we write $\operatorname{Hom}(-,-)$ and $\operatorname{Ext}(-,-)$ for the functors $\operatorname{Hom}_\Lambda(-,-)$ and $\operatorname{Ext}^1_\Lambda(-,-)$, respectively. We fix a minimal injective cogenerator I for $\operatorname{Mod} k$ and write $D = \operatorname{Hom}_k(-, I)$ for the usual *duality* between left and right Λ-modules.

Now fix a Λ-module M. We define $\operatorname{Prod} M$ to be the full subcategory of all direct summands of products of copies of M. The full subcategory formed by all submodules of modules in $\operatorname{Prod} M$ is denoted by $\operatorname{Cogen} M$. The *perpendicular category* of M is by definition the full subcategory ${}^{\perp}M = \{X \in \operatorname{Mod}\Lambda \mid \operatorname{Ext}(X, M) = 0\}$. We call M *selforthogonal* (more precisely, *product-selforthogonal*) if $\operatorname{Ext}(M^{\alpha}, M) = 0$ for every cardinal α. A selforthogonal module M is *maximal selforthogonal* if $\operatorname{Prod} M \subseteq \operatorname{Prod} N$ implies $\operatorname{Prod} M = \operatorname{Prod} N$ for every selforthogonal module N.

Let $\mathcal{C} \subseteq \operatorname{mod}\Lambda$ be a subcategory. Then we denote by $\varinjlim \mathcal{C}$ the full subcategory formed by the filtered colimits of objects in $\mathcal{C}$.

1.1. A correspondence. In this section we establish a correspondence between cotilting modules and certain torsion pairs for $\operatorname{mod}\Lambda$. We start with some preparations.

Lemma 1.1. *Let $\mathcal{X} \subseteq \operatorname{Mod}\Lambda$ be a subcategory which is closed under taking subobjects and filtered colimits. Let $\mathcal{C} = \mathcal{X} \cap \operatorname{mod}\Lambda$. Then $\mathcal{X} = \varinjlim \mathcal{C}$.*

Proof. Clear, since every module is a filtered colimit of its finitely generated submodules. □

Recall that a *torsion pair* $(\mathcal{T}, \mathcal{F})$ for an abelian category $\mathcal{A}$ is a pair of full subcategories such that $\operatorname{Hom}(T, F) = 0$ for all $T \in \mathcal{T}$, $F \in \mathcal{F}$, and every object $X \in \mathcal{A}$ has a subobject $tX \in \mathcal{T}$ with $X/tX \in \mathcal{F}$.

Lemma 1.2. *Let $(\mathcal{T}, \mathcal{F})$ be a torsion pair for $\operatorname{mod}\Lambda$. Then $(\varinjlim \mathcal{T}, \varinjlim \mathcal{F})$ is a torsion pair for $\operatorname{Mod}\Lambda$.*

Proof. See [**9**]. □

We need to recall some more terminology. To this end fix a category $\mathcal{A}$ and some subcategory $\mathcal{X}$. Given an object M in $\mathcal{A}$, a map $X \to M$ is called *right $\mathcal{X}$-approximation* of M provided that X belongs to $\mathcal{X}$ and the induced map $\operatorname{Hom}(X', X) \to \operatorname{Hom}(X', M)$ is surjective for all X' in $\mathcal{X}$. A right approximation $f\colon X \to M$ is *minimal* if every endomorphism $g\colon X \to X$ satisfying $f \circ g = f$ is an isomorphism. The subcategory $\mathcal{X}$ is *contravariantly finite* if every object in $\mathcal{A}$ admits a right $\mathcal{X}$-approximation.

Lemma 1.3. *Let $\mathcal{X} \subseteq \operatorname{Mod}\Lambda$ be a subcategory which is closed under taking subobjects, extensions, products, and filtered colimits. Then every Λ-module M has a minimal right $\mathcal{X}$-approximation $M_{\mathcal{X}} \to M$. The kernel of this approximation is pure-injective. Moreover, $M_{\mathcal{X}}$ is pure-injective provided that M is pure-injective.*

Proof. We use an idea from [**16**, Section 2]. Choose a left $D\mathcal{X}$-approximation $DM \to DX$ which exists by [**16**, Lemma 2.1], and denote by C its cokernel.

Apply the duality and take the pullback along the natural map $M \to D^2M$:

$$\begin{tikzcd}
0 \arrow[r] & DC \arrow[r] \arrow[d, equal] & Y \arrow[r, "f"] \arrow[d] & M \arrow[d] \\
0 \arrow[r] & DC \arrow[r] & D^2X \arrow[r] & D^2M.
\end{tikzcd}$$

Note that Y belongs to $\mathcal{X}$ since $D^2\mathcal{X} \subseteq \mathcal{X}$ by [**16**, Lemma 4.4]. The construction implies that f is a right $\mathcal{X}$-approximation of M. A minimal right $\mathcal{X}$-approximation $M_{\mathcal{X}} \to M$ exists since $\mathcal{X}$ is closed under filtered colimits; see [**11**, p. 207]. It is well-known that any module of the form DX is pure-injective. Therefore the kernel of $M_{\mathcal{X}} \to M$ is pure-injective since it is a direct summand of DC. The natural map $M \to D^2M$ is a pure monomorphism. This implies the map splits if M is pure-injective. In this case $M_{\mathcal{X}}$ is a direct summand of D^2X and therefore pure-injective. □

The next proposition describes a construction for cotilting modules. In [**2**], a similar result is proved under the additional hypothesis that $\mathcal{X}$ is contravariantly finite.

Proposition 1.4. *Let $\mathcal{X} \subseteq \operatorname{Mod}\Lambda$ be a subcategory which is closed under taking subobjects, extensions, products, and filtered colimits. Suppose in addition $\Lambda \in \mathcal{X}$. Then there exists a Λ-module T having the following properties:*

1. *There exists an exact sequence $0 \to T'' \to T' \to Q \to 0$ such that $T = T' \amalg T''$ and Q is an injective cogenerator;*
2. *T is pure-injective and $\operatorname{id} T \leq 1$;*
3. *$\mathcal{X} = \operatorname{Cogen} T = {}^{\perp}T$;*
4. *$\mathcal{X} \cap \mathcal{X}^{\perp} = \operatorname{Prod} T$.*

Proof. (1) We apply Lemma 1.3 to construct T. Choose an injective cogenerator Q and let $0 \to T'' \to T' \to Q \to 0$ be an exact sequence such that $T' \to Q$ is a minimal right $\mathcal{X}$-approximation. Define $T = T' \amalg T''$. Note that $T'' \in \mathcal{X}^{\perp}$ by Wakamatsu's lemma. Thus $T \in \mathcal{X}^{\perp}$.

(2) T is pure-injective by Lemma 1.3. To show $\operatorname{id} T \leq 1$, take an arbitrary Λ-module M and choose an exact sequence $0 \to X_1 \to X_0 \to M \to 0$ with $X_i \in \mathcal{X}$ which exists by Lemma 1.3. Now apply $\operatorname{Hom}(-, T)$ to see that $\operatorname{Ext}^2(M, T) = 0$ since $\operatorname{Ext}(\mathcal{X}, T) = 0$.

(3) Let M be a Λ-module and $M \to Q^{\alpha}$ a monomorphism. The map $M \to Q^{\alpha}$ factors through the map $(T')^{\alpha} \to Q^{\alpha}$ if $M \in {}^{\perp}T$. Therefore ${}^{\perp}T \subseteq \operatorname{Cogen} T$. We have $\operatorname{Cogen} T \subseteq \mathcal{X}$ since T belongs to $\mathcal{X}$ and $\mathcal{X}$ is closed under taking subobjects and products. Finally, $\mathcal{X} \subseteq {}^{\perp}T$ since $T \in \mathcal{X}^{\perp}$.

(4) We have $\operatorname{Prod} T \subseteq \mathcal{X} \cap \mathcal{X}^{\perp}$ since $\mathcal{X}$ and $\mathcal{X}^{\perp}$ are closed under products. Let $M \in \mathcal{X} \cap \mathcal{X}^{\perp}$. Choose a left $\operatorname{Prod} T$-approximation $f\colon M \to T^{\alpha}$ which

is a monomorphism since $\mathcal{X} = \operatorname{Cogen} T$. Now apply $\operatorname{Hom}(-,T)$ to see that $\operatorname{Coker} f$ belongs to ${}^{\perp}T = \mathcal{X}$. Thus $M \to T^{\alpha}$ splits. $\square$

Theorem 1.5. *Let Λ be a left noetherian ring. There exists a bijection between:*

- *Torsion pairs $(\mathcal{T},\mathcal{F})$ for $\operatorname{mod}\Lambda$ such that $\Lambda \in \mathcal{F}$, and*
- *equivalence classes of cotilting modules.*

A cotilting module T corresponding to a torsion pair $(\mathcal{T},\mathcal{F})$ satisfies

$$\operatorname{Prod} T = (\varinjlim \mathcal{F}) \cap (\varinjlim \mathcal{F})^{\perp} \quad \textit{and} \quad {}^{\perp}T \cap \operatorname{mod}\Lambda = \mathcal{F}.$$

Remark 1.6. Suppose Λ is an artin algebra. Then a cotilting module T is equivalent to a finitely generated cotilting module if and only if $\mathcal{F} = {}^{\perp}T \cap \operatorname{mod}\Lambda$ is contravariantly finite in $\operatorname{mod}\Lambda$. This follows from the correspondence for cotilting modules in [**3**].

Proof of Theorem 1.5. Let $(\mathcal{T},\mathcal{F})$ be a torsion pair for $\operatorname{mod}\Lambda$. Then $\mathcal{X} = \varinjlim \mathcal{F}$ is closed under taking subobjects, extensions, products, and filtered colimits by Lemma 1.2. Thus we get a cotilting module T satisfying $\operatorname{Prod} T = \mathcal{X} \cap \mathcal{X}^{\perp}$ by Proposition 1.4.

Conversely, let T be a cotilting module. Then ${}^{\perp}T = \varinjlim \mathcal{F}$ for $\mathcal{F} = {}^{\perp}T \cap \operatorname{mod}\Lambda$ by Lemma 1.1. Thus we obtain a torsion pair $(\mathcal{T},\mathcal{F})$ for $\operatorname{mod}\Lambda$ since $\mathcal{X}$ is a torsion-free class for $\operatorname{Mod}\Lambda$. It is straightforward to check that the maps $(\mathcal{T},\mathcal{F}) \mapsto T$ and $T \mapsto (\mathcal{T},\mathcal{F})$ are mutually inverse. $\square$

1.2. Perpendicular categories. Our next aim is to show that ${}^{\perp}M$ is closed under taking products for a pure-injective module M with $\operatorname{id} M \leq 1$. For this we need to assume that Λ is left artinian. We start with some preparations.

Lemma 1.7. *Let $\mathcal{F} \subseteq \operatorname{mod}\Lambda$ be a subcategory which is closed under taking subobjects and extensions. Let $\mathcal{T}$ be the subcategory formed by all $M \in \operatorname{mod}\Lambda$ satisfying $\operatorname{Hom}(M,\mathcal{F}) = 0$. Then $(\mathcal{T},\mathcal{F})$ is a torsion pair for $\operatorname{mod}\Lambda$.*

Proof. Each finitely generated module M has a minimal submodule tM such that M/tM belongs to $\mathcal{F}$. For this we use that M is artinian and that $\mathcal{F}$ is closed under subobjects. The submodule tM belongs to $\mathcal{T}$ since $\mathcal{F}$ is closed under extensions. $\square$

Each torsion-free class in $\operatorname{Mod}\Lambda$ is closed under taking products. We have therefore the following consequence:

Proposition 1.8. *Let Λ be left artinian. Suppose $\mathcal{X} \subseteq \operatorname{Mod}\Lambda$ is a subcategory which is closed under taking subobjects, extensions, and filtered colimits. Then $\mathcal{X}$ is closed under taking products.*

Proof. Let $\mathcal{F} = \mathcal{X} \cap \operatorname{mod}\Lambda$. Then we have $\mathcal{X} = \varinjlim \mathcal{F}$ by Lemma 1.1, which is a torsion-free class by Lemma 1.7 and 1.2. Thus $\mathcal{X}$ is closed under products. $\square$

Our application about perpendicular categories is based on the following well-known fact:

Lemma 1.9. *Let $\{M_i\}$ be a filtered system of modules and N be a pure-injective module. Then* $\operatorname{Ext}(\varinjlim M_i, N) \cong \varprojlim \operatorname{Ext}(M_i, N)$.

Corollary 1.10. *Let Λ be left artinian. Suppose M is a pure-injective module of injective dimension at most* 1. *Denote by $\mathcal{F}$ the class of finitely generated modules in ${}^{\perp}M$. Then ${}^{\perp}M = \varinjlim \mathcal{F}$. Moreover, ${}^{\perp}M$ is closed under taking products.*

1.3. Complements. Next we discuss when a selforthogonal module T has a complement T' such that $T \amalg T'$ is a cotilting module. This problem has been studied before, for instance in [**1**]. Here we use Corollary 1.10 and get some new result. We need the following well-known lemma which is based on a construction due to Bongartz:

Lemma 1.11. *Let T be a selforthogonal module and Q be an injective cogenerator. Suppose ${}^{\perp}T$ is closed under taking products. Then there exists a module T' having the following properties:*

(1) *There exists an exact sequence $0 \to T^{\alpha} \to T' \to Q \to 0$ for some cardinal α.*
(2) *$T \amalg T'$ is selforthogonal and ${}^{\perp}T = {}^{\perp}(T \amalg T')$.*
(3) *If T is pure-injective, then T' is pure-injective.*

Proof. We sketch the proof for the convenience of the reader. The crucial idea is that of a *universal extension* [**4**, **8**]. That is, take the product of all exact sequences in $\operatorname{Ext}(Q, T)$ and let $\alpha = \operatorname{card}(\operatorname{Ext}(Q, T))$. Now take the pullback along the codiagonal map $Q \to Q^{\alpha}$, to obtain a sequence

$$0 \longrightarrow T^{\alpha} \longrightarrow T' \longrightarrow Q \longrightarrow 0.$$

Clearly, ${}^{\perp}T = {}^{\perp}(T \amalg T')$. The construction of the universal extension implies $\operatorname{Ext}(T', T) = 0$. Thus $T \amalg T'$ is selforthogonal.

The last assertion follows from the following fact: A module M satisfying (C1)-(C3) in the definition of a cotilting module is pure-injective if and only if ${}^{\perp}M$ is closed under taking filtered colimits and pure submodules; see Proposition 5.7 in [**16**]. □

Corollary 1.12. *Let Λ be left artinian. Suppose T is a pure-injective module of injective dimension at most* 1 *satisfying* $\operatorname{Ext}(T, T) = 0$. *Then there exists a module T' such that $T \amalg T'$ is a cotilting module.*

1.4. A generalization. The correspondence in Theorem 1.5 can be generalized as follows: Let $\mathcal{A}$ be a locally noetherian Grothendieck category, that is, $\mathcal{A}$ is a Grothendieck category and has a generating set of noetherian objects. Recall that a set $\mathcal{G}$ of objects *generates* an additive category $\mathcal{C}$ if for every nonzero map f in $\mathcal{C}$ we have $\operatorname{Hom}(G, f) \neq 0$ for some $G \in \mathcal{G}$. Next we recall that an object X in $\mathcal{A}$ is *pure-injective* if every pure exact sequence $\varepsilon\colon 0 \to X \to Y \to Z \to 0$ splits. Here, ε is *pure exact* if $\operatorname{Hom}(A,\varepsilon)$ is an exact sequence for every noetherian object A in $\mathcal{A}$. Finally, we say that an object T in $\mathcal{A}$ is a *cotilting object* if T satisfies the conditions (C1)-(C4).

Theorem 1.13. *Let $\mathcal{A}$ be a locally noetherian Grothendieck category and denote by* $\operatorname{noeth}\mathcal{A}$ *the full subcategory formed by all noetherian objects. There exists a bijection between:*

- *Torsion pairs $(\mathcal{T},\mathcal{F})$ for* $\operatorname{noeth}\mathcal{A}$ *such that $\mathcal{F}$ generates* $\operatorname{noeth}\mathcal{A}$*, and*
- *equivalence classes of cotilting objects.*

A cotilting object T corresponding to a torsion pair $(\mathcal{T},\mathcal{F})$ satisfies

$$\operatorname{Prod} T = (\varinjlim \mathcal{F}) \cap (\varinjlim \mathcal{F})^{\perp} \quad \text{and} \quad {}^{\perp}T \cap \operatorname{noeth}\mathcal{A} = \mathcal{F}.$$

Proof. The proof is essentially the same as that of Theorem 1.5. However, it remains to explain the analogue of the duality D between left and right modules which is used in the proof of Lemma 1.3. To this end let $\mathcal{C} = \operatorname{noeth}\mathcal{A}$ and denote by $(\mathcal{C}^{\mathrm{op}},\mathrm{Ab})$ the category of additive functors $F\colon \mathcal{C}^{\mathrm{op}} \to \mathrm{Ab}$. The functor

$$\mathcal{A} \longrightarrow (\mathcal{C}^{\mathrm{op}},\mathrm{Ab}), \quad X \mapsto \operatorname{Hom}(-,X)|_{\mathcal{C}},$$

identifies $\mathcal{A}$ with the full subcategory of left exact functors $\mathcal{C}^{\mathrm{op}} \to \mathrm{Ab}$; see [**12**]. The duality $D = \operatorname{Hom}_{\mathbb{Z}}(-,\mathbb{Q}/\mathbb{Z})$ induces a duality between $(\mathcal{C}^{\mathrm{op}},\mathrm{Ab})$ and $(\mathcal{C},\mathrm{Ab})$ via the assignment $F \mapsto D\circ F$. It is straightforward to check that the proof of Lemma 1.3 works with this duality. $\square$

2. Selforthogonal modules.

Throughout the rest of this paper we assume that Λ is a tame hereditary algebra. We refer to Ringel's Rome notes [**18**] for basic facts about the representation theory of tame hereditary algebras. In this section we discuss pure-injective Λ-modules which are selforthogonal. A basic tool are the following Auslander-Reiten formulas:

$$D\operatorname{Ext}(X,M) \cong \operatorname{Hom}(M,\tau X) \text{ and } \operatorname{Ext}(M,X) \cong D\operatorname{Hom}(\tau^{-1}X,M)$$

which are valid for every Λ-module M and every finitely generated module X; see [**10**]. Here, τX denotes the *dual of transpose* of X.

2.1. Pure-injectives. Let $\operatorname{Ind}\Lambda$ denote the set of isoclasses of indecomposable pure-injective Λ-modules and let $\operatorname{ind}\Lambda$ be the set of finitely generated objects in $\operatorname{Ind}\Lambda$. The classification of indecomposable pure-injectives is well-known:

$$\operatorname{Ind}\Lambda = \operatorname{ind}\Lambda \cup \{S_\infty \mid S \in \mathbb{P}\} \cup \{S_{-\infty} \mid S \in \mathbb{P}\} \cup \{G\}.$$

The category of finitely generated modules decomposes into three subcategories: The preprojectives $\mathcal{P}$, the preinjectives $\mathcal{I}$, and the regular modules

$\mathcal{R}$. The category $\mathcal{R}$ is abelian and a Λ-module is called *quasi-simple* if it is a simple object in $\mathcal{R}$. The following picture indicates the direction of nonzero maps between indecomposable pure-injectives:

$$\begin{tikzcd}
 & & G \arrow[dr] & & \\
 & S_{-\infty} \arrow[ur] \arrow[dr] & & S_{\infty} \arrow[dr] & \\
\mathcal{P} \arrow[ur] \arrow[rr] & & \mathcal{R} \arrow[ur] \arrow[rr] & & \mathcal{I}.
\end{tikzcd}$$

Proposition 2.1. *Let M be a pure-injective module. Then there exists a family of indecomposable pure-injective modules $(M_i)_{i\in I}$ such that M is a pure-injective envelope of $\coprod_{i\in I} M_i$. The M_i are unique up to isomorphism; they are, up to isomorphism, precisely the indecomposable direct summands of M.*

Proof. See Proposition 8.33 and Theorem 8.53 in [**15**]. □

The description of the pure-injectives has some useful consequences.

Lemma 2.2. *Let M be a pure-injective module. then ${}^{\perp}M = \bigcap_N {}^{\perp}N$ where N runs through all indecomposable direct summands of M.*

Proof. This follows immediately from Proposition 2.1. Write $M = E(\coprod_i M_i)$ as the pure-injective envelope of a coproduct of indecomposables. Then we have

$$\bigcap_i {}^{\perp}M_i = {}^{\perp}\left(\prod_i M_i\right) \subseteq {}^{\perp}M \subseteq \bigcap_i {}^{\perp}M_i$$

since M is a direct summand of $\prod_i M_i$. □

The following result simplifies the description of the pure-injective Λ-modules which are selforthogonal:

Corollary 2.3. *For a pure-injective module M the following are equivalent:*

(1) $\mathrm{Ext}(M, M) = 0$;
(2) $\mathrm{Ext}(M', M'') = 0$ *for all indecomposable direct summands* M', M'' *of* M;
(3) M *is selforthognal, that is,* $\mathrm{Ext}(M^\alpha, M) = 0$ *for every cardinal* α.

Proof. We know from Corollary 1.10 that ${}^\perp M$ is closed under taking products. Writing $M = E(\coprod_i M_i)$ as the pure-injective envelope of a coproduct of indecomposables, we see that ${}^\perp M$ contains M since M is a direct summand of $\prod_i M_i$. $\square$

2.2. Extensions between indecomposables. We construct non-split extensions between some indecomposable pure-injectives.

Lemma 2.4. *Let S be a quasi-simple module, then there is a* (*non-split*) *exact sequence*

$$0 \longrightarrow (\tau S)_{-\infty} \longrightarrow \coprod G \longrightarrow S_\infty \longrightarrow 0. \tag{2.1}$$

Proof. Let $r \in \mathbb{N}$ be the minimal number such that $\tau^r S = S$. There is a system of exact sequences

$$\begin{tikzcd}
0 \arrow[r] & S_r \arrow[r] \arrow[d, equal] & S_{r+1} \arrow[r] \arrow[d] & S_1 \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & S_r \arrow[r] \arrow[d, equal] & S_{r+2} \arrow[r] \arrow[d] & S_2 \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & S_r \arrow[r] \arrow[d, equal] & S_{r+3} \arrow[r] \arrow[d] & S_3 \arrow[r] \arrow[d] & 0 \\
 & \vdots & \vdots & \vdots &
\end{tikzcd}$$

which induces an exact sequence $0 \to S_r \to S_\infty \xrightarrow{f} S_\infty \to 0$ with $\operatorname{Ker} f^n = S_{nr}$ for all $n \in \mathbb{N}$. It is well-known that the inverse limit of the system

$$\cdots \xrightarrow{f} S_\infty \xrightarrow{f} S_\infty \xrightarrow{f} S_\infty$$

is a coproduct of copies of G. Using the formula $S_{nr} = (\tau S)_{-nr}$, we obtain the following system of exact sequences:

$$\begin{tikzcd}
 & \vdots \arrow[d] & \vdots \arrow[d, "f"] & \vdots \arrow[d, equal] & \\
0 \arrow[r] & (\tau S)_{-3r} \arrow[r] \arrow[d] & S_\infty \arrow[r, "f^3"] \arrow[d, "f"] & S_\infty \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & (\tau S)_{-2r} \arrow[r] \arrow[d] & S_\infty \arrow[r, "f^2"] \arrow[d, "f"] & S_\infty \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & (\tau S)_{-1r} \arrow[r] & S_\infty \arrow[r, "f^1"] & S_\infty \arrow[r] & 0.
\end{tikzcd}$$

Take inverse limits to obtain the required sequence. It is exact because the system is surjective, and thus the Mittag-Leffler condition (see e.g., [**19**]) is satisfied. □

Lemma 2.5. *Let M be an indecomposable pure-injective module which is not finitely generated.*

(1) $\operatorname{Ext}(M, Q) \neq 0$ *for every nonzero preprojective module Q.*
(1) $\operatorname{Ext}(J, M) \neq 0$ *for every nonzero preinjective module J.*

Proof. First we assume $M = S_\infty$. It is well-known that for any preprojective module Q there is a nonzero map to a module in every tube. Thus there is a nonzero map $\tau^{-1}Q \to S'_r \to S'_\infty$ for some quasi-simple $S' \sim S$. The Auslander-Reiten formula implies $\operatorname{Ext}(S'_\infty, Q) \neq 0$. Suppose $S = \tau^n S'$. There exists an exact sequence $0 \to S_n \to S_\infty \to S'_\infty \to 0$ which induces a monomorphism $\operatorname{Ext}(S'_\infty, Q) \to \operatorname{Ext}(S_\infty, Q)$ since $\operatorname{Hom}(S_n, Q) = 0$. Thus $\operatorname{Ext}(S_\infty, Q) \neq 0$.

Now let J be preinjective and suppose there is a nonzero map $S'_r \to \tau J$ for some quasi-simple $S' \sim S$. We find an epimorphism $S_n \to S'_r$ for some $n \in \mathbb{N}$. Thus $\operatorname{Ext}(J, S_n) \neq 0$ by the Auslander-Reiten formula. The exact sequence $0 \to S_n \to S_\infty \to (\tau^{-n}S)_\infty \to 0$ induces a monomorphism $\operatorname{Ext}(J, S_n) \to \operatorname{Ext}(J, S_\infty)$ since $\operatorname{Hom}(J, (\tau^{-n}S)_\infty) = 0$. Thus $\operatorname{Ext}(J, S_\infty) \neq 0$.

The statements for $S_{-\infty}$ follow from the statements for S_∞ by duality. More precisely, we have $S_{-\infty} = D((DS)_\infty)$, and we combine this fact with the formula $\operatorname{Ext}(M, DN) \cong \operatorname{Ext}(N, DM)$ which is valid for all modules M and N.

It remains to consider the generic module G. This case reduces to the previous cases since the exact sequence (2.1) induces monomorphisms $\operatorname{Ext}(S_\infty, Q) \to \operatorname{Ext}(\coprod G, Q)$ and $\operatorname{Ext}(J, (\tau S)_{-\infty}) \to \operatorname{Ext}(J, \coprod G)$. □

Next we compute the perpendicular categories for Prüfer modules and the generic module.

Lemma 2.6. *Let M be the generic or a Prüfer module. Then* ${}^{\perp}M = \varinjlim(\mathcal{P} \cup \mathcal{R})$.

Proof. There is a torsion pair $(\mathcal{I}, \mathcal{P} \cup \mathcal{R})$ for the category of finitely generated Λ-modules. We have that $\mathcal{P}$ and $\mathcal{R}$ belong to ${}^{\perp}M$, whereas $\mathcal{I} \cap {}^{\perp}M = 0$ by Lemma 2.5. Thus ${}^{\perp}M = \varinjlim(\mathcal{P} \cup \mathcal{R})$ by Corollary 1.10. $\square$

We need to know when $\operatorname{Ext}(M, N)$ vanishes for indecomposable pure-injective modules M and N.

Lemma 2.7. *Let M and N be indecomposable pure-injective Λ-modules which are not finitely generated. Then $\operatorname{Ext}(M, N) \neq 0$ if and only if there are quasi-simples $S \sim S'$ in $\mathbb{P}$ such that $M \cong S_\infty$ and $N \cong S'_{-\infty}$.*

Proof. Suppose first that $S \sim S'$ are quasi-simples in $\mathbb{P}$. The exact sequence (2.1) induces an injective map

$$\operatorname{Hom}(\tau S_\infty, S'_\infty) \longrightarrow \operatorname{Ext}(\tau S_\infty, \tau S'_{-\infty}) \xrightarrow{\sim} \operatorname{Ext}(S_\infty, S'_{-\infty})$$

since $\operatorname{Hom}(S_\infty, G) = 0$. Thus $\operatorname{Ext}(S_\infty, S'_{-\infty}) \neq 0$ follows since there is a nonzero map $\tau S_\infty \to S'_\infty$.

Now suppose $N = G$ or $N = S_\infty$ for some quasi-simple S. Then $\operatorname{Ext}(M, N) = 0$ by Lemma 2.6 since $S'_\infty \in \varinjlim \mathcal{R}$ and $S'_{-\infty} \in \varinjlim \mathcal{P}$ for each quasi-simple S'. For $M = G$, use the exact sequence (2.1).

Finally, let $N = S_{-\infty}$. We have $\operatorname{Ext}(S'_{-\infty}, N) = 0$ for each quasi-simple S' since $\mathcal{P} \subseteq {}^{\perp}N$. Now assume $S' \not\sim S$. Then $\operatorname{Ext}(\tau^i S', S_{-\infty}) = 0$ for all $i \geq 0$ by Lemma 4 in [**10**]. Therefore $\operatorname{Ext}(S'_n, S_{-\infty}) = 0$ for all $n \geq 1$. Thus $\operatorname{Ext}(S'_\infty, N) \cong \varprojlim \operatorname{Ext}(S'_n, N) = 0$. Using again the exact sequence (2.1), we see that $\operatorname{Ext}(G, N) = 0$. $\square$

We need also a statement which involves finitely generated modules.

Lemma 2.8. *Let S and S' be quasi-simples such that $S \not\sim S'$. Then $\operatorname{Ext}(S_n, S'_m) = 0$ for all $n, m \in \mathbb{N} \cup \{-\infty, \infty\}$.*

Proof. Observe that for each $n \in \mathbb{N}$ the module S_n is a submodule of S_∞ and a quotient of $S_{-\infty}$. Now the assertion follows from Lemma 2.7 since $\operatorname{Ext}(-,-)$ is right exact. $\square$

3. Cotilting modules.

3.1. A characterization. In this section we prove our main result about cotilting modules for tame hereditary algebras. First we show that cotilting modules are precisely the maximal objects among all selforthogonal modules. This observation leads to some useful characterizations.

Proposition 3.1. *The following are equivalent for a pure-injective Λ-module T:*

(1) *T is a cotilting module;*
(2) *T is maximal selforthogonal.*

Proof. (1) $\Rightarrow$ (2): First observe that that ${}^{\perp}T \subseteq \operatorname{Cogen} T$. To see this, denote by Q an injective cogenerator for $\operatorname{Mod}\Lambda$ and let X be a module in ${}^{\perp}T$. There is a monomorphism $X \to Q^{\alpha}$. Consider the pullback

$$\begin{tikzcd} 0 \arrow[r] & T' \arrow[r] \arrow[d, equal] & Y \arrow[r] \arrow[d] & X \arrow[r] \arrow[d] & 0 \\ 0 \arrow[r] & T' \arrow[r] & T'' \arrow[r] & Q^{\alpha} \arrow[r] & 0 \end{tikzcd}$$

with T', T'' in $\operatorname{Prod} T$. The upper sequence splits and we obtain a monomorphism $X \to T''$.

Now assume X is a module such that $T \amalg X$ is selforthogonal. Since X is in $\operatorname{Cogen} T$, we can choose a left $\operatorname{Prod} T$-approximation of X, to get an exact sequence

$$\xi\colon 0 \longrightarrow X \longrightarrow T' \longrightarrow Y \longrightarrow 0$$

that remains exact when one applies $\operatorname{Hom}(-, T)$ to it. Therefore one gets that Y is also in ${}^{\perp}T$. Then there is an exact sequence

$$0 \longrightarrow Y \longrightarrow T'' \longrightarrow Z \longrightarrow 0$$

with $T'' \in \operatorname{Prod} T$. We apply $\operatorname{Hom}(-, X)$ to this sequence and obtain $\operatorname{Ext}(Y, X) = 0$ since $\operatorname{id} X \leq 1$. Thus ξ splits. This means X belongs to $\operatorname{Prod} T$, so T is maximal selforthogonal.

(2) $\Rightarrow$ (1): First observe that ${}^{\perp}T$ is closed under taking products by Corollary 1.10. Now apply Lemma 1.11. $\square$

Having shown that cotilting modules are maximal selforthogonal modules, we can apply our results about selforthogonal modules. First we separate the finitely generated cotilting modules from those having indecomposable summands which are not finitely generated.

Proposition 3.2. *Let T be a cotilting module with a finitely generated preprojective or preinjective direct summand. Then T is equivalent to a finitely generated cotilting module.*

Proof. Apply Lemma 2.5. $\square$

Next we introduce for each tube $\mathcal{T}$ in the category of regular modules the $\mathcal{T}$-component of a module and use this to characterize the cotilting modules.

Let $\mathbb{X} = \mathbb{P}/{\sim}$ be the set of equivalence classes of quasi-simples. Note that we obtain a decomposition $\mathcal{R} = \coprod_{\sigma \in \mathbb{X}} \mathcal{T}_{\sigma}$ of $\mathcal{R}$ into connected abelian categories, where each quasi-simple S belongs to $\mathcal{T}_{\sigma}$ for $\sigma = [S]$. The categories $\mathcal{T}_{\sigma}$ are usually called *tubes* and $\operatorname{card} \sigma$ is called the *rank* of the tube

$\mathcal{T}_\sigma$. We say that a pure-injective module M *belongs* to a tube $\mathcal{T}$ if every indecomposable direct summand of M is of the form S_n with $S \in \mathcal{T}$ and $n \in \mathbb{N} \cup \{-\infty, \infty\}$. The subcategory formed by all modules belonging to $\mathcal{T}$ is denoted by $\overline{\mathcal{T}}$. Given a pure-injective module M and a tube $\mathcal{T}$, we define the $\mathcal{T}$-component $M_\mathcal{T}$ of M to be a maximal direct summand of M belonging to $\overline{\mathcal{T}}$. Note that $M_\mathcal{T}$ is unique up to isomorphism.

Proposition 3.3. *Let T be a pure-injective module without a nonzero finitely generated preprojective or preinjective direct summand. Then the following are equivalent:*

(1) *T is a cotilting module;*
(2) *each $T_\mathcal{T}$ is maximal selforthogonal among all modules in $\overline{\mathcal{T}}$.*

Proof. (1) ⇒ (2): We know from Proposition 3.1 that T is maximal selforthogonal. In particular, each $T_\mathcal{T}$ is selforthogonal. Assume there is a tube $\mathcal{T}$ such that $T_\mathcal{T}$ is not maximal selforthogonal. Then there exists X in $\overline{\mathcal{T}}$ with X not in $\operatorname{Prod} T_\mathcal{T}$ such that $T_\mathcal{T} \amalg X$ is selforthogonal. The module X has no extensions with the direct summands of T belonging to different tubes, by Lemma 2.8. It follows from Corollary 2.3 that $T \amalg X$ is selforthogonal in $\operatorname{Mod} \Lambda$, since T has no preprojective or preinjective summands. This is a contradiction.

(2) ⇒ (1): First observe that T has an indecomposable direct summand which is not finitely generated. This follows from the fact that for each tube $\mathcal{T}$ of rank 1, $T_\mathcal{T}$ has no finitely generated summands.

Next we need to show that T is selforthogonal. By Corollary 2.3, it is sufficient to show $\operatorname{Ext}(T', T'') = 0$ for indecomposable direct summands T', T'' of T. If T' and T'' belong to different tubes, then it follows from Lemma 2.8 that $\operatorname{Ext}(T', T'') = 0$. If they belong to the same tube, then $\operatorname{Ext}(T', T'') = 0$ by the assumption on $T_\mathcal{T}$.

Assume T is not a cotilting module. Using Lemma 1.11, we find a pure-injective module X such that $T \amalg X$ is a cotilting module. Choose an indecomposable direct summand X' of X which does not belong to $\operatorname{Prod} T$. It follows from Proposition 3.2 that X' is neither preprojective nor preinjective since T has an indecomposable summand which is not finitely generated. Thus X' belongs to a tube $\mathcal{T}$. It follows that $T_\mathcal{T} \amalg X'$ is a selforthogonal module in $\overline{\mathcal{T}}$. A contradiction. □

Remark 3.4. Given a tube $\mathcal{T}$, one can show that the modules in $\overline{\mathcal{T}}$ are precisely the filtered limits and the filtered colimits of modules in $\mathcal{T}$.

3.2. Selforthogonal modules in tubes. Fix a tube $\mathcal{T}$ of rank r. We say that a pure-injective module M belonging to $\mathcal{T}$ is of *Prüfer type* if M has no adic module as a direct summand. Analogously, M is of *adic type* if M has no Prüfer module as a direct summand. Next we define a bijection $M \mapsto M^\vee$

between the modules of Prüfer and adic type in $\overline{\mathcal{T}}$. To this end number the quasi-simples $S(1),\dots,S(r)$ in $\mathcal{T}$ such that $\tau^j S(i) = S(i+j)$ for $i,j \in \mathbb{Z}_r$. If $M = S(i)_n$ is indecomposable, let $M^\vee = S(-i)_{-n}$. If $M = E(\coprod_i M_i)$ with all M_i indecomposable, let $M^\vee = E(\coprod_i M_i^\vee)$.

Lemma 3.5. *The map $M \mapsto M^\vee$ induces a bijection between the modules of Prüfer type and the modules of adic type in $\overline{\mathcal{T}}$. Moreover, M is selforthogonal if and only if $M^\vee$ is selforthogonal.*

Proof. It follows immediately from the definition that $M^{\vee\vee} \cong M$ for all $M \in \overline{\mathcal{T}}$. Therefore $M \mapsto M^\vee$ induces a bijection between the modules of Prüfer and adic type.

The rest follows from Lemma 3.6 given below. In addition, we use Corollary 2.3 which says that a pure-injective module M is selfortogonal if $\operatorname{Ext}(M', M'') = 0$ for all indecomposable direct summands M', M'' of M. $\square$

Lemma 3.6. *Let $i,j \in \mathbb{Z}_r$ and $n,m \in \mathbb{N} \cup \{\infty\}$. Then* $\operatorname{Ext}(S(i)_n, S(j)_m) = 0$ *if and only if* $\operatorname{Ext}(S(-j)_{-m}, S(-i)_{-n}) = 0$.

Proof. First observe that $S_{-n} = D((DS)_n)$ for any quasi-simple S. The assertion follows from the formula $\operatorname{Ext}(DM, DN) \cong \operatorname{Ext}(N, D^2M)$ which is valid for all modules M and N; it is a consequence of the formula $\operatorname{Ext}(M, DN) \cong \operatorname{Ext}(N, DM)$. In addition, one uses that $D^2(S_\infty)$ is a coproduct of copies of S_∞. $\square$

Each tube $\mathcal{T}$ gives rise to a locally finite Grothendieck category $\varinjlim \mathcal{T}$. Using Theorem 1.13, we can classify its cotilting objects and get a description of the selforthogonal modules of Prüfer type. This leads to the following result which is taken from [**6**]:

Proposition 3.7. *Let M be a selforthogonal module of Prüfer type belonging to a tube $\mathcal{T}$. Then M is maximal among all selforthogonal modules in $\overline{\mathcal{T}}$ if and only if the number of indecomposable non-isomorphic direct summands of M equals the rank of $\mathcal{T}$.*

Proposition 3.7 extends to arbitrary selforthogonal modules belonging to tubes.

Corollary 3.8. *Let M be a selforthogonal module belonging to a tube $\mathcal{T}$. Then M is maximal among all selforthogonal modules in $\overline{\mathcal{T}}$ if and only if the number of indecomposable non-isomorphic direct summands of M equals the rank of $\mathcal{T}$.*

Proof. Each selforthogonal module is of Prüfer or of adic type. This follows from Lemma 2.7. Now use Proposition 3.7 and Lemma 3.5. $\square$

3.3. The main theorem. We are now in the position to prove our main result about tame hereditary algebras.

Theorem 3.9. *Let Λ be a tame hereditary algebra and let T be a pure-injective Λ-module.*

(1) *Suppose all indecomposable direct summands are finitely generated. Then T is a cotilting module if and only if the number of non-isomorphic indecomposable direct summands of T equals the number of simple Λ-modules and* $\operatorname{Ext}(T',T'')=0$ *for all* $T',T''\in\operatorname{indec}T$.

(2) *Suppose there is an indecomposable direct summand which is not finitely generated. Then T is a cotilting module if and only if the following hold:*
   - *Each indecomposable direct summand of T is either generic or of the form S_n for some $S\in\mathbb{P}$ and some $n\in\mathbb{N}\cup\{-\infty,\infty\}$.*
   - *For each tube $\mathcal{T}$, let $I_{\mathcal{T}}$ be the set of non-isomorphic indecomposable direct summands of T which are of the form S_n for some $n\in\mathbb{N}\cup\{-\infty,\infty\}$ and some quasi-simple $S\in\mathcal{T}$. Then* $\operatorname{card}I_{\mathcal{T}}$ *equals the rank of $\mathcal{T}$ and* $\operatorname{Ext}(T',T'')=0$ *for all* $T',T''\in I_{\mathcal{T}}$.

(3) *Two cotilting modules T_1 and T_2 are equivalent if and only if* $\operatorname{indec}(T_1\amalg G)=\operatorname{indec}(T_2\amalg G)$.

Proof. (1) This is a well-known result from [**4**].

(2) Assume first T is a cotilting module with a summand which is not finitely generated. Then there are no preprojective or preinjective direct summands, by Proposition 3.2. It follows that all direct summands of T are either generic or of the form S_n for some quasi-simple $S\in\mathbb{P}$ and some $n\in\mathbb{N}\cup\{-\infty,\infty\}$. Now fix a tube $\mathcal{T}$ and consider the $\mathcal{T}$-component $T_{\mathcal{T}}$ of T. Proposition 3.3 implies that $T_{\mathcal{T}}$ is maximal selforthogonal among all modules in $\overline{\mathcal{T}}$. Thus $\operatorname{card}I_{\mathcal{T}}$ equals the rank of $\mathcal{T}$ by Corollary 3.8.

Suppose now T is a module such that each direct summand of T is either generic or of the form S_n for some quasi-simple $S\in\mathbb{P}$ and some $n\in\mathbb{N}\cup\{-\infty,\infty\}$. Suppose also that $\operatorname{card}I_{\mathcal{T}}$ equals the rank of $\mathcal{T}$ and $\operatorname{Ext}(T',T'')=0$ for all $T',T''\in I_{\mathcal{T}}$. It follows from Corollary 2.3 that $T_{\mathcal{T}}$ is selforthogonal for each tube $\mathcal{T}$. Moreover, Corollary 3.8 implies that $T_{\mathcal{T}}$ is maximal selforthogonal among all modules in $\overline{\mathcal{T}}$. Thus T is a cotilting module by Proposition 3.3.

(3) Recall the well-known fact that any finitely generated indecomposable module X arises as a direct summand of a product $\prod_i M_i$ if X is a direct summand of one of the M_i. It is therefore sufficient to consider statement (3) in case T_1 and T_2 are both not finitely generated. All one needs to show is that the generic module G is the only indecomposable module not in $\operatorname{indec}T$ that can occur in $\operatorname{Prod}T$, for a cotilting module T which is not finitely generated. Let $T_1\sim T_2$ and suppose M is in $\operatorname{indec}T_2$, but not in $\operatorname{indec}T_1$. Then $T=M\amalg T_1$ is cotilting. Assume M is not generic. Then

there is a tube $\mathcal{T}$ such that M is in $\overline{\mathcal{T}}$. It follows that $T_{\mathcal{T}}$ is not selforthogonal. This contradiction shows M is generic. $\square$

For the Kronecker algebra, there are no finitely generated selforthogonal regular modules. We obtain therefore the following description of all cotilting modules, up to equivalence:

Corollary 3.10. *Let T be a cotilting module over a Kronecker algebra. Then T is equivalent to one of the following:*

- *A finitely generated preinjective cotilting module,*
- *a finitely generated preprojective cotilting module,*
- *a cotilting module of the form*

$$\left(\coprod_{S\in\mathbb{P}'} S_\infty\right) \amalg E\left(\coprod_{S\in\mathbb{P}''} S_{-\infty}\right)$$

for some partition $\mathbb{P}' \cup \mathbb{P}''$ of $\mathbb{P}$.

4. The maximal selforthogonal objects in tubes.

Let us complete the characterization of the cotilting modules over a tame hereditary algebra. To this end we describe the maximal selforthogonal modules belonging to a tube. In [**6**], a complete combinatorial description of the maximal selforthogonal objects of Prüfer type is given. Using Lemma 3.5, one also gets a description of the maximal selforthogonal objects of adic type.

Let $\mathcal{T}$ be a tube of rank r in $\operatorname{Mod}\Lambda$. Let $S(1), S(2), \ldots, S(r)$ be the quasi-simples, ordered such that $\tau^j S(i) = S(i+j)$ for $i, j \in \mathbb{Z}_r$. The *ray vector* of a module M in $\overline{\mathcal{T}}$ is an r-tuple $(a_1, a_2, \ldots, a_r)$ of nonnegative integers such that M has exactly a_i indecomposable direct summands with quasi-socle $S(i)$.

Proposition 4.1. *Let $\mathcal{T}$ be a tube of rank r. For each r-tuple $(a_1, a_2, \ldots, a_r)$ of nonnegative integers with $\sum_i a_i = r$, there is exactly one maximal selforthogonal object T of Prüfer type in $\overline{\mathcal{T}}$ with this m-tuple as its ray vector.*

We refer to [**6**] for the proof; it also gives an algorithm how to find a maximal selforthogonal object with a given ray-vector.

Given a module M in $\overline{\mathcal{T}}$, the *coray vector* of M is an r-tuple $(b_1, b_2, \ldots, b_r)$, where b_i is the number of indecomposable direct summands in M with quasi-top isomorphic to $S(i)$.

Proposition 4.2. *Let $\mathcal{T}$ be a tube of rank r. For each r-tuple $(b_1, b_2, \ldots, b_r)$ of nonnegative integers with $\sum_i b_i = m$, there is exactly one maximal selforthogonal object T of adic type in $\overline{\mathcal{T}}$, with this r-tuple as its coray vector.*

Proof. The correspondence $M \mapsto M^\vee$ assigns to a maximal selforthogonal object of Prüfer type and with ray vector $(a_1, a_2, \ldots, a_r)$, a maximal selforthogonal object of adic type and with coray vector $(a_{r-1}, a_{r-2}, \ldots, a_r)$. □

Acknowledgments. We would like to thank Helmut Lenzing, Claus M. Ringel and Øyvind Solberg for helpful comments and ideas. In addition, we are grateful to Lidia Angeleri Hügel for pointing out a mistake in one of our arguments.

References

[1] H.L. Angeleri and F.-U. Coelho, *Infinitely generated complements to partial tilting modules*, Math. Proc. Cambridge Phil. Soc., **132** (2002), 89-96, CMP 1 866 326.

[2] H.L. Angeleri, A. Tonolo and J. Trlifaj, *Tilting preenvelopes and cotilting precovers*, Algebr. Represent. Theory, **4** (2001), 155-170, MR 2002e:16010, Zbl 0999.16007.

[3] M. Auslander and I. Reiten, *Applications of contravariantly finite subcategories*, Adv. Math., **86** (1991), 111-152, MR 92e:16009, Zbl 0774.16006.

[4] K. Bongartz, *Tilted algebras*, in 'Representations of algebras (Puebla, 1980),' Lecture Notes in Math., **903**, Springer, Berlin-New York, 1981, 26-38, MR 83g:16053, Zbl 0478.16025.

[5] S. Brenner and M.C.R. Butler, *Generalizations of the Bernstein-Gelfand-Ponomarev reflection functors*, in 'Representation theory, II' (Proc. Second Internat. Conf., Carleton Univ., Ottawa, Ont., 1979), Lecture Notes in Math., **832**, Springer, Berlin-New York, 1980, 103-169, MR 83e:16031, Zbl 0446.16031.

[6] A.B. Buan and H. Krause, *Tilting and cotilting for quivers of type $\widetilde{A}_n$*. Preprint, 2002.

[7] R. Colpi, G. D'Este and A. Tonolo, *Quasi-tilting modules and counter equivalences*, J. Algebra, **191** (1997), 461-494, MR 99g:16008, Zbl 0876.16004.

[8] R. Colpi, A. Tonolo and J. Trlifaj, *Partial cotilting modules and the lattice induced by them*, Comm. Algebra, **25** (1997), 3225-3237, MR 98i:16003, Zbl 0893.16017.

[9] W.W. Crawley-Boevey, *Locally finitely presented additive categories*, Comm. Algebra, **22** (1994), 1644-1674, MR 95h:18009, Zbl 0798.18006.

[10] ———, *Infinite-dimensional modules in the representation theory of finite-dimensional algebras*, CMS Conf. Proc., **23** (1998), 29-54, MR 99m:16016, Zbl 0920.16007.

[11] E. Enochs, *Injective and flat covers, envelopes and resolvents*, Israel J. Math., **39** (1981), 189-209, MR 83a:16031, Zbl 0464.16019.

[12] G. Gabriel, *Des catégories abéliennes*, Bull. Soc. Math. France, **90** (1962), 323-448, MR 38 #1144, Zbl 0201.35602.

[13] W. Geigle, *The Krull-Gabriel dimension of the representation theory of a tame hereditary Artin algebra and applications to the structure of exact sequences*, Manuscripta Math., **54** (1985), 83-106, MR 87e:16061, Zbl 0593.16022.

[14] R. Göbel and J. Trlifaj, *Cotilting and a hierarchy of cotorsion groups*, J. Algebra, **224** (2000), 110-122, MR 2001a:20100, Zbl 0947.20036.

[15] C. Jensen and H. Lenzing, *Model-Theoretic Algebra with Particular Emphasis on Fields, Rings, Modules*, Algebra, Logic and Applications, **2**, Gordon and Breach Science Publishers, New York, 1989, MR 91m:03038, Zbl 0728.03026.

[16] H. Krause and Ø. Solberg, *Filtering modules of finite projective dimension*, Forum Math., to appear.

[17] H. Lenzing, *Homological transfer from finitely presented to infinite modules*, in 'Abelian group theory (Honolulu, Hawaii, 1983),' Lecture Notes in Math., **1006**, Springer, Berlin, 1983, 734-761, MR 85f:16034, Zbl 0529.16016.

[18] C.M. Ringel, *Infinite-dimensional representations of finite-dimensional hereditary algebras*, Symposia Mathematica, Vol. XXIII (Conf. Abelian Groups and their Relationship to the Theory of Modules, INDAM, Rome, 1977), Academic Press, London-New York, 1979, 321-412, MR 81i:16032 , Zbl 0429.16022.

[19] J. Rotman, *An Introduction to Homological Algebra*, Pure and Applied Mathematics, **85**, Academic Press, Inc., New York-London, 1979, MR 80k:18001, Zbl 0441.18018.

Received April 8, 2002 and revised November 6, 2002.

Institutt for Matematiske fag
NTNU
N-7491 Trondheim
Norway
E-mail address: aslakb@math.ntnu.no

Department of Pure Mathematics
University of Leeds
Leeds LS2 9JT
United Kingdom
E-mail address: henning@maths.leeds.ac.uk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

THE CONSTANT CURVATURE PROPERTY OF THE WU INVARIANT METRIC

C.K. Cheung and Kang-Tae Kim

We investigate the property of the Wu invariant metric on a certain class of psuedoconvex domains. We show that the Wu invariant Hermitian metric, which in general behaves as nicely as the Kobayashi metric under holomorphic mappings, enjoys the complex hyperbolic curvature property in such cases. Namely, the Wu invariant metric is Kähler and has constant negative holomorphic curvature in a neighborhood of the spherical boundary points for a large class of domains in $\mathbb{C}^n$.

1. Introduction and the main theorems.

The invariant metrics including the Carathéodory, Kobayashi, Bergman and Kähler-Einstein metrics (to name only a few) have played an important role in the complex function theory, complex geometry and many related areas of mathematical research. Each invariant metric provides distinct merits and characteristics, and hence its own unique contributions. Among them, the Carathéodory and Kobayashi metrics are exploited mostly because of the distance decreasing property, that all holomorphic mappings are distance decreasing with respect to them. Thus they provide a bridge between the complex function theory and geometry. On the other hand, they are almost never Hermitian, in some sense making them harder to be applied by the differential geometric methods. The Bergman and Kähler-Einstein metrics are Hermitian (in fact Kählerian), and hence they offer wide range of applications in geometric research. However, the distance decreasing property for general holomorphic mappings essentially fail with them. (See Wu [**7**] for more detailed discussions on this point.) Along such a circle of reasons, it was natural for H. Wu to introduce the metric that is Hermitian and distance-decreasing for all holomorphic mappings. We call this metric *the Wu metric* throughout this paper.

This article focuses upon the curvature properties of the Wu metric. As one considers the distance decreasing properties of holomorphic mappings with respect to the given Hermitian/Kählerian metrics, well-known differential geometric generalizations of Schwarz's lemma (see [**1**], [**8**] for instance) show that the negativity of curvature can imply the distance decreasing property for holomorphic mappings. On the other hand, the Kobayashi metric, while not being Hermitian in general, always enjoys the distance decreasing property. Thus, it was natural for Kobayashi to ask the following question:

Question. Do all Kobayashi hyperbolic manifolds admit a Hermitian metric of negative holomorphic curvature?

Here, by a Kobayashi hyperbolic manifold we mean a complex manifold on which the Kobayashi distance is positive definite. The terminology *holomorphic curvature* stands for the holomorphic sectional curvature.

Notice that the Hermitian metrics which do not possess distance decreasing property (up to a positive constant multiplier) cannot answer the above question. It appeals to us that hopes, if any, lie with the Wu metric. This has inspired us to consider the curvature behavior of the Wu metric in various cases.

Indeed, the main focus of this article is centered around the curvature analysis of the Wu metric on the bounded pseudoconvex domains. In the preceding analysis of the Wu metric on the Thullen domains ([**2**] and [**3**]), we have observed several totally unexpected aspects of its metric and curvature behavior. When one considers the Wu metric on the Thullen domain $E_2 = \{(z,w) \in \mathbb{C}^2 \mid |z|^2 + |w|^4 < 1\}$ for instance, the Wu metric is real analytic except at points belonging to the subsets

$$Z = \{(z,0) \in \mathbb{C}^2 \mid |z| < 1\} \text{ and } M = \{(z,w) \in E_4 \mid |z|^2 + 2|w|^4 = 1\}.$$

Surprisingly, it is C^1 but not C^2 at points of M. Moreover, the Wu metric is not Kähler "inside" M with a variable holomorphic curvature, whereas it is Kähler "outside" M with a constant holomorphic curvature. It seems that such a "singularity set" for the Wu metric is originated by the weak pseudoconvexity of boundary points of the Thullen domain E_4. On the other hand, the results of [**2**] and [**3**] involve much computation, even for the points outside M. As several experts pointed out after [**2**] and [**3**] appeared, a simple and straightforward geometric explication of the constant curvature property is desirable. The main point of this paper is exactly focusing upon this; we give a more general and geometric explanation for the boundary constant curvature property of the Wu metric. As a consequence of our new arguments, we show that the constant curvature property is not only possessed by the complex two dimensional Thullen domains but also shared by much broader a collection of domains that map locally properly onto the ball, for instance. The main results are as follows:

Theorem 1. *Let Ω be a domain obtained by intersecting the ball B^n with an open set. Denote the boundaries of Ω and B^n by $\partial\Omega$ and ∂B^n respectively. There exists a neighborhood V of $\partial\Omega \cap \partial B^n$ such that the Wu metric of Ω is Kähler with constant negative holomorphic curvautre in $V \cap \Omega$.*

Theorem 2. *Let* $\widetilde{E} = \{(z_1, \ldots, z_n) \in \mathbb{C}^n : |z_1|^{2m_1} + \cdots + |z_n|^{2m_n} < 1\}$, *where* $m_1, \ldots, m_n$ *are integers* ≥ 2. *If* p *is a strongly pseudoconvex boundary point of* $\widetilde{E}$, *then there exists a neighborhood* V_p *of* p *such that the Wu metric of* $\widetilde{E}$ *is Kähler with constant negative holomorphic curvautre in* $V_p \cap \widetilde{E}$.

Note that Theorem 2 provides a generalization of the complex hyperbolicity results of [**2**]. Furthermore, its generalization is immediate from the proof; see the remark at the end of Section 5.

Remark. We would like to point out that Theorem 1 cannot be completely localized. The domain $\{(z, w) \in \mathbb{C}^2 \mid |z|^2 + \sqrt{|w|} < 1\}$ is locally biholomorphic with the ball in a neighborhood of each smooth boundary point. However, the holomorphic curvature of its Wu metric is nowhere constant and nowhere Kähler. See [**3**] for details.

2. Wu metric and the minimal ellipsoid.

Let us first recall the definition of the Kobayashi-Royden invariant metric. Let B^1 be the unit disk in $\mathbb{C}$ and M be a complex manifold. Denote by $H(B^1, M)$ the set of all holomorphic maps from B^1 to M. The Kobayashi-Royden metric of M is defined by

$$k_M(p, v) = \inf \Big\{ |u| : u \in \mathbb{C} \text{ and } df(0)(u) = v \\ \text{for some } f \in H(B^1, M) \text{ with } f(0) = p \Big\}$$

for all $p \in M$ and $v \in T_p(M)$. Here we identify $T_0(B^1)$ with $\mathbb{C}$ and $|\ |$ is the euclidean norm.

In [**7**], Wu introduced a new invariant *Hermitian* metric. The construction and properties of the Wu metric are discussed in [**2**], [**5**] and [**7**]. While we point the readers to these articles for a detailed introduction, we include here a brief description of the Wu metric for the Kobayashi hyperbolic manifolds. If M is a Kobayashi hyperbolic complex manifold, then its Wu metric h_M and the Kobayashi-Royden infinitesimal metric k_M are related as shown in the following proposition.

First let us fix $x \in M$. Let α be a positive definite Hermitian inner product on $T_x M$. Denote its unit ball by

$$B_\alpha = \{v \in T_x M : \alpha(v, v) \leq 1\}.$$

Likewise, consider

$$I_x^M = \{v \in T_x M : k_M(x; v) \leq 1\}$$

which is commonly called the *Kobayashi indicatrix of* M *at* x. Then, we have:

Proposition 1 (Wu [**7**]). *The Wu metric h_x on T_xM satisfies that:*
(1) $I_x^M \subseteq B_{h_x}$, *and*
(2) volume of $B_{h_x} \leq$ volume of B_g *for every* $g \in \mathcal{P}_x$ *that satisfies* $I_x^M \subseteq B_g$,

where the volume is measured by any Hermitian inner product on T_xM and $\mathcal{P}_x$ is the set of all positive definite Hermitian inner products on T_xM.

In other words, in each tangent space of a Kobayashi hyperbolic manifold, the unit ball of the Wu metric is the *unique* complex ellipsoid centered at the origin with smallest possible volume among those containing the Kobayashi indicatrix. (See also [**4**] for the uniqueness property.) In this article, we call this ellipsoid the *minimal ellipsoid.*

Now we present:

Lemma 1. *Suppose D is a subset of $\overline{B}^n$. Let $L_k = \{(z_1, \ldots, z_n) \in \overline{B}^n : |z_1|^2 \leq |z_k|^2$ and $z_j = 0$ for $j \neq k\}$, where $k = 2, \ldots, n$. If $\bigcup_{k=2}^n L_k \subset D$, then the minimal ellipsoid of D is $\overline{B}^n$.*

This lemma states that if a domain containing a "sufficiently large portion" of the boundary of the complex unit ball ∂B^n, then its corresponding minimal ellipsoid is the closed ball $\overline{B}^n$.

Proof. Let $E = \{\sum_i a_i|z_i|^2 + 2\mathrm{Re}\sum_{i\neq j} a_{ij}z_i\overline{z}_j \leq 1\}$, for $a_i > 0$ with $i, j = 1, \ldots, n$, be the minimal ellipsoid that contains $\bigcup_{k=2}^n L_k$. Assume that E is not the closed ball $\overline{B}^n$. Notice that E is not invariant under the action of the group

$$T = \{(z_1, \ldots, z_n) \mapsto (e^{\sqrt{-1}\theta_1}, \ldots, e^{\sqrt{-1}\theta_n}) \mid \theta_i \in \mathbb{R}, i = 1, \ldots, n\}$$

unless $a_{ij} = 0$. Since the set $\bigcup_{k=2}^n L_k$ is invariant under the action by T above, and since this action is volume preserving, the uniqueness of the minimal ellipsoid implies that $a_{ij} = 0$. The ellipsoid is thus of the form $E = \{\sum_i a_i|z_i|^2 \leq 1\}$. Since $L_k \subset E$, we have $a_1 \geq 1$ and $a_k \leq 1$ for $k = 2, \ldots, n$. The volume of E is less than that of the B^n implies that $a_1a_2 \cdot \cdots \cdot a_n > 1$. Hence there exists an integer j' between 2 to n such that $a_1a_{j'} > 1$. Without loss of generality, we can assume that $j' = 2$ with $a_1a_2 > 1$. The previous statements can now be reduced to a two dimensional case as follows:

$$E' = \{(z_1, z_2) : a_1|z_1|^2 + a_2|z_2|^2 \leq 1\} \supset L_2' = \{(z_1, z_2) \in \overline{B}^2 : |z_1|^2 \leq |z_2|^2\}.$$

This can further imply that in the first quadrant of $\mathbb{R}^2$,

$$E'' = \{a_1x + a_2y \leq 1\} \supset L_2'' = \{x + y \leq 1, x \leq y\}.$$

But a simple geometrical argument can easily conclude that if E'' contains L_2'' then $a_1a_2 \leq 1$. This contradicts the assumption that E is not the closed unit ball. Since $\bigcup_{k=2}^{n} L_k \subset D \subset \overline{B}^n$, it follows that the minimal ellipsoid of D is also the closed unit ball. $\square$

3. The extremal discs in B^n.

Let B^1 be the unit disc in $\mathbb{C}$, and Ω be a domain in $\mathbb{C}^n$. A holomorphic map $f : B^1 \to \Omega$ is called an *extremal mapping* at $p \in \Omega$ in the complex direction v, if it realizes the Kobayashi-Royden metric at p along v. The image, $f(B^1)$, is then called an *extremal disc* in Ω. See [**6**].

In this section, we study the behavior of certain extremal discs in B^n for points near the boundary.

Consider the point $\widetilde{a} = (a, 0, \dots, 0) \in B^n$. The automorphism $\phi : B^n \to B^n$ defined by

$$\phi(w_1, \dots, w_n) \mapsto \left(\frac{w_1 + a}{1 + \overline{a} w_1}, \frac{\sqrt{1 - |a|^2} w_2}{1 + \overline{a} w_1}, \dots, \frac{\sqrt{1 - |a|^2} w_n}{1 + \overline{a} w_1} \right)$$

sends $(0, \dots, 0)$ to $\widetilde{a}$. At $(0, \dots, 0)$, the extremal disc in the direction $(X_1, \dots, X_n) \in \mathbb{C}^n$ is given by the complex line $F : B^1 \to B^n$ with $F(\zeta) = \zeta\,(X_1, \dots, X_n)$. (Here for simplicity we have assumed that $(X_1, \dots, X_n) \in S$, the unit sphere in $\mathbb{C}^n$.) Since the extremal disc is invariant under biholomorphism, the image $\phi \circ F(B^1)$ becomes the extremal disc at $\widetilde{a}$ in the direction of the new vector

$$(\phi \circ F)'(0) = \left((1 - |a|^2) X_1, \sqrt{1 - |a|^2} X_2, \dots, \sqrt{1 - |a|^2} X_n \right),$$

and the corresponding extremal mapping is

$$(\phi \circ F)(\zeta) = \left(\frac{\zeta X_1 + a}{1 + \overline{a} \zeta X_1}, \frac{\sqrt{1 - |a|^2} \zeta X_2}{1 + \overline{a} \zeta X_1}, \dots, \frac{\sqrt{1 - |a|^2} \zeta X_n}{1 + \overline{a} \zeta X_1} \right).$$

If, in addition, $|X_1|^2 \le |X_k|^2$, for a certain $k > 1$ and $X_j = 0$ for every $j \ne k$, we have $|X_1| \le \frac{1}{\sqrt{2}}$. The properties of Möbius transformations then implies that the extremal map $(\phi \circ F)(\zeta)$ converges uniformly to $(1, 0, \dots, 0)$ as $a \to 1$. In other words, we have the following lemma:

Lemma 2. *Given a neighborhood U at $(1, 0, \dots, 0)$, there exists a constant $\epsilon > 0$ such that if $|1 - a| < \epsilon$, then for any $(X_1, X_2, \dots, X_n) \in S$ with the property that for certain k between 1 and n, $|X_1| \le |X_k|$ while $X_j = 0$ for all $j \ne k$, the extremal disc at $(a, 0, \dots, 0)$ in the complex direction*

$$\left((1 - |a|^2) X_1, \sqrt{1 - |a|^2} X_2, \dots, \sqrt{1 - |a|^2} X_n \right)$$

lies inside $U \cap B^n$.

4. Proof of Theorem 1.

Suppose Ω be a domain obtained by intersecting B^n with an open set W. Let $z \in \Omega = B^n \cap W$. Rotating the axis if necessary, we may assume without loss of generality that $z = (a, 0, \dots, 0)$. By Lemma 2 , if z is close enough to the boundary of B^n, the extremal discs of B^n at z along certain complex directions lie inside Ω. Since $\Omega \subset B^n$, the definition of the Kobayashi metric implies that these extremal discs (for B^n) are also extremal for the domain Ω. To state this fact more precisely in terms of Kobayashi indicatrix, we have

$$\widetilde{L}_k \subset I_z^\Omega \subset I_z^{B^n}, \text{ for } k = 2, \dots, n$$

where

$$\widetilde{L}_k = \Big((1-|a|^2)Y_1, \sqrt{1-|a|^2}Y_2, \dots, \sqrt{1-|a|^2}Y_n \Big)$$

with $(Y_1, \dots, Y_n) \in S$ and $|Y_1| \le |Y_k|$, while $Y_j = 0$, for every $j \ne k$. Now, consider the automorphism $\phi : B^n \to B^n$ defined in Section 3, that sends $0 = (0, \dots, 0)$ to $z = (a, 0, \dots, 0)$. By identifying the tangent spaces with $\mathbb{C}^n$, we then have

$$(d\phi_0)^{-1}(\widetilde{L}_k) \subset (d\phi_0)^{-1}(I_z^\Omega) \subset (d\phi_0)^{-1}(I_z^{B^n}).$$

Notice that $(d\phi_0)^{-1}(I_z^{B^n}) = \overline{B}^n$ and

$$(d\phi_0)^{-1}(\widetilde{L}_k) = \{(w_1, \dots, w_n) \in S : |w_1|^2 \le |w_k|^2 \text{ and } w_j = 0 \text{ for } j \ne k\}.$$

Lemma 1 can now be used to conclude that the minimal ellipsoid of $(d\phi_0)^{-1}(I_z^\Omega)$ is the same as the closed unit ball. Since the minimal ellipsoids are determined solely by the linear structure, it follows that

$$\text{minimal ellipsoid of } I_z^\Omega = I_z^{B^n}.$$

This implies that the Wu metric of Ω at this point z coincides with the Poincare-Bergman metric of the ball. Notice that if $p \in \partial\Omega \backslash \partial W \subset \partial B^n$, then there exists an open neighborhood V_p of p such that the above argument holds for every $z \in V_p$. This completes the proof of Theorem 1. □

5. Proof of Theorem 2.

Let m_1 and m_2 are integers ≥ 2,

$$\begin{aligned} E_{m_1 m_2} &= \{(z_1, z_2) \in \mathbb{C}^2 : |z_1|^{2m_1} + |z_2|^{2m_2} < 1\}, \\ \Sigma &= \{(z_1, 0) \in \mathbb{C}^2 : |z_1| \le 1\} \cup \{(0, z_2) \in \mathbb{C}^2 : |z_2| \le 1\}, \\ \partial\Sigma &= \{(z_1, 0) : |z_1| = 1\} \cup \{(0, z_2) : |z_2| = 1\}. \end{aligned}$$

In this section, we investigate the Wu metric h_E of $E = E_{m_1 m_2}$ near $p \in \partial E \backslash \Sigma$. Consider $\Phi(z_1, z_2) = (z_1^{m_1}, z_2^{m_2})$. Then Φ induces a proper holomorphic mapping from $E_{m_1 m_2}$ onto B^2 with its branch set Σ. Take an open neighborhood $U = U_p$, say, of p so that $\Phi|_U$ is 1-1. Let $z \in U \cap E$. By

Theorem 1, if z is close enough to the boundary point p, then the Wu metric of $U' = \Phi(U) \cap B^2$ at $\Phi(z)$ is the same as the Poincare-Bergman metric of the ball. Denote the Kobayashi indicatrices of U' and B^2 by $I^{U'}_{\Phi(z)}$ and $I^{B^2}_{\Phi(z)}$, respectively. Then we have

$$\text{minimal ellipsoid of } I^{U'}_{\Phi(z)} = I^{B^2}_{\Phi(z)}.$$

Since the Wu metric is invariant under biholomorphism, it follows that

$$\begin{aligned}\text{minimal ellipsoid of } I^{U\cap E}_{z} &= \text{minimal ellipsoid of } (d\Phi_z)^{-1} I^{U'}_{\Phi(z)} \\ &= (d\Phi_z)^{-1} I^{B^2}_{\Phi(z)}.\end{aligned}$$

On the other hand, Φ, being holomorphic, is a distance decreasing mapping with respect to the Kobayashi metrics of E and B^2. Consequently, we have

$$I^{E}_{z} \subset (d\Phi_z)^{-1} I^{B^2}_{\Phi(z)}.$$

Recall that $U \cap E$ is a subset of E. Thus we have

$$\text{minimal ellipsoid of } I^{U\cap E}_{z} \subset \text{minimal ellipsoid of } I^{E}_{z}.$$

Therefore, we can conclude that minimal ellipsoid of I^{E}_{z} equals to $(d\Phi_z)^{-1} I^{B^2}_{\Phi(z)}$. In particular, the Wu metric of E at z is given by

$$h_E(z) = (\Phi^* h_{B^2})(\Phi(z)).$$

Notice that there exists an open neighborhood $V_p \subset U$ of p such that the above argument holds for every $z \in V_p$. Now we have

$$h_E = \Phi^* h_{B^2} \text{ at every } z \in V_p.$$

It is evident to see that this argument is also valid for the n dimensional case. Therefore, the proof of Theorem 2 is now complete. □

Remark. The same proof can also be used to show the following: Let $\Omega = \{(z_1, z_2) \in \mathbb{C}^2 : |f_1(z_1, z_2)|^2 + |f_2(z_1, z_2)|^2 < 1\}$, where $(f_1, f_2) : \Omega \to B^2$ is a proper holomorphic mapping. Then, the set of boundary points away from branch locus of the proper holomorphic mapping admits an open neighborhood on which the Wu metric is complex hyperbolic.

Acknowledgments. It is our pleasure to acknowledge our indebtedness to H. Wu for his encouragements, interest and many helpful discussions with the authors.

References

[1] L. Ahlfors, *An extension of Schwarz's Lemma*, Trans. Amer. Math. Soc., **43** (1938), 359-364, CMP 1 501 949, Zbl 0018.41002.

[2] C.K. Cheung and K.T. Kim, *Analysis of the Wu metric* I: *The case of convex Thullen domain*, Trans. Amer. Math. Soc., **358** (1996), 1429-1457, MR 96i:32026, Zbl 0872.32017.

[3] ______, *Analysis of the Wu metric* II: *The case of non-convex Thullen domain*, Proc. Amer. Math. Soc., **125** (1997), 1131-1142, MR 97f:32029, Zbl 0872.32018.

[4] F. John, *Extremum problems with inequalities as subsidiary conditions*, Studies and Essays Presented to Richard Courant on his 60th Birthday, January 8, 1948, Interscience, New York, 1948, 187-204, MR 10,719b, Zbl 0034.10503.

[5] K. Kim, *The Wu metric and minimal ellipsoids*, Proceedings of the 3rd Pacific Rim Geometry Conference, (J. Choe ed.), Monogr. Geom. Topology, **25**, International Press, Cambridge, 1998, 121-138, MR 2001f:32017.

[6] L. Lempert, *La métrique de Kobayashi et la representation des domains sur la boule*, Bull. Soc. Math. France, **109** (1981), 427-474, MR 84d:32036, Zbl 0492.32025.

[7] H. Wu, *Old and new invariant metrics*, Several complex variables: Proc. of Mittag-Leffler Inst., 1987-88 (J.E. Fornaess ed.), Math. Notes, Princeton Univ. Press, **38** (1993), 640-682, MR 94a:32038, Zbl 0773.32017.

[8] S.-T. Yau, *A general Schwarz lemma for Kähler manifolds*, American J. Math., **100** (1978), 197-203, MR 58 #6370, Zbl 0424.53040.

Received March 9, 2000. Research of the second named author is supported in part by POSTECH Special Fund and KOSEF Grant 981-0104-018-2 and 1999-2-102-003-5 of The Republic of Korea.

Department of Mathematics
Boston College
Chestnut Hill, MA 02467
E-mail address: ck.cheung@bc.edu

Department of Mathematics
Pohang University of Science and Technology
Pohang 790-784
The Republic of Korea
E-mail address: kimkt@postech.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

CONJUGACY AND COUNTEREXAMPLE IN RANDOM ITERATION

Mark Comerford

We consider counterexamples in the field of random iteration to two well-known theorems of classical complex dynamics - Sullivan's non-wandering theorem and the classification of periodic Fatou components. Random iteration which was first introduced by Fornaess and Sibony (1991) is a generalization of standard complex dynamics where instead of considering iterates of a fixed rational function, one allows the mappings to vary at each stage of the iterative process. In this setting one can produce oscillatory behaviour of a type forbidden in classical rational iteration. The technique of the proof requires us to extend the classical notion of conjugacy between dynamical systems to random iteration and we prove some basic results concerning conjugacy in this setting.

1. Introduction.

We consider a sequence of rational functions $\{R_n\}_{n=1}^{\infty} = \{R_1, R_2, R_3, \dots\}$ of some fixed degree $d \geq 2$. Let $Q_n(z)$ be the composition of the first n of these functions in the natural order, i.e.,

$$Q_n = R_n \circ R_{n-1} \circ \cdots\cdots \circ R_2 \circ R_1.$$

We will also be interested in the compositions

$$Q_{m,n} = R_n \circ R_{n-1} \circ \cdots\cdots \circ R_{m+2} \circ R_{m+1}.$$

We now define the *Fatou set* $\mathcal{F}$ for such a sequence of rational functions as

$$\mathcal{F} = \{z \in \overline{\mathbb{C}} : \{Q_n\}_{n=1}^{\infty} \text{ is a normal family on some neighbourhood of } z\}$$

and the *Julia set* $\mathcal{J}$ is then simply the complement of the Fatou set in $\overline{\mathbb{C}}$. These definitions were first introduced by Fornaess and Sibony in [**9**]. Note that, if $\{R_n\}_{n=1}^{\infty}$ is a constant sequence, then these definitions coincide with the standard ones. One important consequence of this definition of Julia and Fatou sets is that we can formulate an analogue of the principle of complete invariance in standard rational iteration. In order to do this, we need to introduce the following terminology:

Let $\{R_n\}_{n=1}^{\infty}$ be as above and for any fixed $n \geq 0$, let us define the *n-th iterated Julia set* $\mathcal{J}_n$ and *n-th iterated Fatou set* $\mathcal{F}_n$ to be the Julia and

Fatou sets for the sequence $\{R_{n+1}, R_{n+2}, R_{n+3}, \dots\}$ which we obtain from our original sequence simply by deleting the first n members. Note that with these definitions, $\mathcal{J}_0 = \mathcal{J}$ and $\mathcal{F}_0 = \mathcal{F}$. We then have the following:

Theorem 1.1. *For any $m < n \in \mathbb{Z}_+$, $Q_{m,n}(\mathcal{J}_m) = \mathcal{J}_n$ and $Q_{m,n}(\mathcal{F}_m) = \mathcal{F}_n$, with Fatou components of $\mathcal{F}_m$ being mapped surjectively onto those of $\mathcal{F}_n$.*

The proof is a straightforward adaptation of the standard classical proof.

The notation introduced above can also be extended to sets and points. For a set U which we introduce at stage m and for $n > m$, we set $U_n = Q_{m,n}(U)$ and for a point x which is introduced at stage m, we set $x_n = Q_{m,n}(x)$.

It turns out that the above situation using rational functions is somewhat too general for proving significant results. In fact, even if one restricts oneself to sequences of polynomials, one obtains pathologies which show that this situation is still too far from traditional complex dynamics to develop a meaningful theory [**4**, **6**]. The most natural restriction one can probably make was introduced by Fornaess and Sibony [**9**] who considered sequences of monic polynomials with uniformly bounded coefficients, i.e., sequences of the form

$$R_n(z) = P_n(z) = z^d + a_{d-1,n}z^{d-1} + \cdots\cdots + a_{1,n}z + a_{0,n}$$

and where we can find some $M \geq 0$ such that $|a_{i,n}| \leq M$ for $0 \leq i \leq d-1$ and all $n \geq 1$. From now on, we shall call such sequences *bounded monic polynomial sequences.*

We will also be interested in the more general setting where we consider sequences of polynomials which need no longer be monic but where we still retain some degree of control over the leading coefficients. More specifically we will consider sequences of the form

$$P_n(z) = a_{d,n}z^d + a_{d-1,n}z^{d-1} + \cdots\cdots + a_{1,n}z + a_{0,n}$$

where as before we can find some $M \geq 0$ such that $|a_{i,n}| \leq M$ for $0 \leq i \leq d-1$ and all $n \geq 1$ and we can also find $K \geq 1$ such that $1/K \leq |a_{d,n}| \leq K$ for all $n \geq 1$. From now on, we shall call such sequences *bounded polynomial sequences.* This definition clearly contains the previous one as a special case and one of the advantages of both definitions is that we can find some radius R depending only on the coefficient bounds M, K above so that for any sequence $\{P_n\}_{n=1}^{\infty}$ as above, it is easy to see that

$$|Q_n(z)| \to \infty \qquad \text{as} \quad n \to \infty, \quad |z| > R$$

which shows in particular that as for classical polynomial Julia sets, there will be a basin at infinity $\mathcal{A}_\infty$ on which all points escape to infinity under iteration. Such a radius will be called an *escape radius* for the coefficient bound M and one can employ the maximum principle to show that $\mathcal{A}_\infty$ is completely invariant just as in the classical case. The complement of the basin at infinity $\mathcal{A}_\infty$ is called the *filled Julia set* $\mathcal{K}$. As in the classical case, it is simply the set of points whose orbits remain bounded under iteration and also as in the classical case, it follows by Montel's theorem that $\partial\mathcal{K} = \mathcal{J}$. Clearly $\mathcal{K} \subset D(0, R)$ where R is an escape radius for $\{P_n\}_{n=1}^{\infty}$ and $D(0, R)$ denotes the closed disk of radius R about 0.

Using monic sequences, one can construct the Green's function with pole at infinity for the outer domain $\mathcal{A}_\infty$ by analogy with the classical result as was done by Fornaess and Sibony [**9**]. The Julia set is perfect, regular, and has logarithmic capacity one. The proofs still work for general bounded sequences, only one no longer has such a nice formula for the capacity. However, bounded sequences allow one greater freedom in choosing polynomials for special counterexamples as we shall see later. We now turn to stating the main result of this paper. Classically, we have the following two well-known theorems from complex dynamics:

Theorem 1.2 (Classification of Periodic Fatou Components)**.** *Let R be a rational function and let U be a* (*classical*) *Fatou component for R which is periodic, i.e., $R^{\circ n}(U) = U$ for some $n \geq 1$. Then we have one of the following four possibilities for U:*

1. *U contains a point of an attracting or superattracting cycle.*
2. *U is a basin of a parabolic periodic point lying on ∂U.*
3. *U is a Siegel disk.*
4. *U is a Herman ring.*

Theorem 1.3 (Sullivan)**.** *Let R be a rational function and let U be a* (*classical*) *Fatou component for R. Then U is eventually periodic.*

The first of these results immediately implies that for a periodic Fatou component, all normal limit functions on that component are either constant or univalent. If we combine this with Sullivan's non-wandering theorem, we see that for any given Fatou component the normal limit functions will be either all constant or all nonconstant. A second simple consequence of the non-wandering theorem is that for a given Fatou component, the diameters of the iterates of that component must eventually stabilize in view of the eventual periodicity. For random iteration, however, neither of these results are true as the following theorem shows:

Theorem 1.4. *There exists a bounded monic sequence of polynomials for which there is a Fatou component V with the following properties:*

1. *There are both constant and nonconstant normal limit functions on V.*
2. $\limsup_{n\to\infty} \operatorname{diam}(V_n) > 0$ *but* $\liminf_{n\to\infty} \operatorname{diam}(V_n) = 0$.

This result therefore both gives a direct counterexample to an immediate consequence of Sullivan's theorem and to a simple consequence of Sullivan's theorem and the classification of periodic Fatou components. In addition, the same argument yields a sequence with a Fatou component which satisfies both 1. and 2. For a further counterexample to a simple classical consequence of these results together with the Fatou-Shishikura-Epstein bound on the number of non-repelling cycles, see [**7**]. Finally, we remark that for the situation of iteration with an entire function, examples of wandering domains exist and even examples of wandering domains on which all the iterates are univalent [**8**].

In order to be able to say that there is a bounded monic polynomial sequence with the properties we need, we will need to develop and make precise the notion of conjugacy between random sequences of polynomials and the next section of this paper is devoted to this aim. We will first prove the result for a bounded non-monic sequence of polynomials and then argue by conjugacy that there is a monic sequence with the desired properties.

2. Analytic conjugacy.

We start by recalling the classical definition of analytic conjugacy between polynomials. Two polynomials P^1 and P^2 are conjugate on $\overline{\mathbb{C}}$ if there exists an affine linear mapping $\varphi(z) = \alpha z + \beta$ such that $\varphi \circ P^1 \circ \varphi^{-1} = P^2$.

Our version of this for random iteration is as follows: We start by considering two sequences $\{P_n\}_{n=1}^{\infty}$, $\{\widetilde{P}_n\}_{n=1}^{\infty}$ of polynomials of degree ≥ 2 acting on $\overline{\mathbb{C}}$. We say that two such sequences are conjugate if there exists a sequence of affine linear mappings $\varphi_n(z) = \alpha_n z + \beta_n$ such that $\varphi_n \circ P_n \circ \varphi_{n-1}^{-1} = \widetilde{P}_n$ for every $n \geq 1$. In order to be able to make meaningful comparisons between different sequences, we need the mappings and their inverses to form an equicontinuous family, i.e., there must exist $K \geq 1$, $M > 0$ such that $1/K \leq |\alpha_n| \leq K$, $|\beta_n| \leq M$ for all $n \geq 0$. We note that this definition has appeared earlier in the literature, for example in the paper of Kolyada and Snoha [**10**] where they make use of it in dealing with topological entropy for sequences of mappings of a compact topological space. The following result is immediate from the definitions:

Proposition 2.1. *Let $\{P_n\}_{n=1}^{\infty}$ and $\{\widetilde{P}_n\}_{n=1}^{\infty}$ be two bounded polynomial sequences which are analytically conjugate in the random sense above. Then for any $n \geq 0$, x is in the nth iterated Fatou set $\mathcal{F}_n$ for $\{P_n\}_{n=1}^{\infty}$ if and only if $\varphi_n(x)$ is in the nth iterated Fatou set $\widetilde{\mathcal{F}}_n$ for $\{\widetilde{P}_n\}_{n=1}^{\infty}$. Also U is a Fatou component of $\mathcal{F}_n$ if and only if $\varphi_n(U)$ is a Fatou component of $\widetilde{\mathcal{F}}_n$.*

What is the relationship between classical and random conjugacy in this situation? Clearly any classical conjugacy also gives a conjugacy in the random sense but is the converse true? The answer to this question is not in general. However, in the generic case, we will see that not only are two polynomials which are conjugate in the random sense classically conjugate but also that every random conjugacy between them must in fact also be a classical conjugacy.

Let P^1 and P^2 be two polynomials of the same degree $d \geq 2$ which are conjugate in the random sense via a sequence of affine linear mappings $\varphi_n = \alpha_n z + \beta_n$, $n \geq 0$ and where the constants α_n and β_n are bounded as in the definition. By considering the fact that these mappings must map the barycentre of the set of critical points for P^1 to that of the critical points of P^2, and by classically conjugating with suitable translations, one finds that we may take $\beta_n = 0$ for every $n \geq 0$.

So let us assume from now on that $\varphi_n = \alpha_n z$ for every $n \geq 0$. Let us now turn our attention to the zeros of P^1 and P^2. Each mapping $\alpha_n z$ must map the zeros of P^1 onto those of P^2 and its inverse must map those of P^2 onto those of P^1. Clearly, we may assume that after conjugating P^2 with a rotation and a dilation if necessary, that the zeros of P^1 and those of P^2 do in fact coincide. Let us for now assume that there are zeros of P^1 and P^2 which do not lie at 0. It then follows that we can find $1 \leq q \leq d$ and $r \geq 0$, $s \geq 1$ with $qs + r = d$ together with nonzero constants $a_1, a_2, \ldots, a_s$ and c_1, c_2 so that

$$P^i(z) = c_i z^r \prod_{i=1}^{s} (z^q - a_i), \qquad i = 1, 2.$$

The number q is the degree of rotational symmetry of this set of zeros about 0 and if there is no nontrivial symmetry of this kind, this is equivalent to having $q = 1$.

It follows easily from this that we must have $\alpha_n = e^{2\pi i p_n/q}$ where p_n is a sequence of integers all of which can be taken from among the finite set $\{0, 1, 2, \ldots, q-1\}$. From this we see that $c_2/c_1 = e^{2\pi i p/q}$ for some fixed $0 \leq p < q$. The equation relating p_n and p_{n+1} is therefore

$$p_{n+1} = p + r p_n \qquad \text{in} \quad \mathbb{Z}_q$$

and each initial choice of p_0 gives a potentially different random conjugacy between P^1 and P^2. One observes at this point that the above equation shows that the mappings in any conjugacy are in fact determined by a discrete classical dynamical system on $\mathbb{Z}_q$ given by the mapping $f(i) = p + ri$, $i \in \mathbb{Z}_q$. Hence, finding out what types of conjugacies are permitted between P^1 and P^2 is the same as knowing all possible types of orbit for every element of $\mathbb{Z}_q$ under this dynamical system.

Conjugacies which correspond to fixed points of f will be referred to as *fixed conjugacies* and clearly coincide with classical conjugacies. Periodic orbits of points in $\mathbb{Z}_q$ give rise to conjugacies which we shall refer to as *periodic conjugacies*. On the other hand, since $\mathbb{Z}_q$ is a finite set, every orbit is preperiodic and so we can talk of *preperiodic* and *strictly preperiodic conjugacies*. With this terminology in hand one can deduce the following:

1. P^1 and P^2 are classically conjugate if and only if $r-1$ divides p in $\mathbb{Z}_q$ and all conjugacies between P^1 and P^2 are classical if and only if $p=0$ and $r=1$, both equations being over $\mathbb{Z}_q$.
2. There are periodic conjugacies of period $k \geq 1$ if and only if there exist solutions of the equation
$$(r^k-1)a+(r^{k-1}+r^{k-2}+\cdots+r+1)p=0$$
which are not solutions of the same equation for some smaller value of k (including $k=1$).
3. There are strictly preperiodic conjugacies if and only if $q>1$ and r divides 0 in $\mathbb{Z}_q$.

This finishes the case when there are nonzero roots of P^1 and P^2. We now consider the case when all the zeros of P^1 and P^2 are at 0. In this case, by classically conjugating both polynomials if needed, we may assume that $P^1(z)=P^2(z)=z^d$. Here P^1 and P^2 are obviously classically conjugate but there are other types of conjugacy as well. For example, the sequence of mappings $\{\omega z, \overline{\omega} z, \omega z, \overline{\omega} z, \dots\}$ where $\omega = e^{2\pi i/(d+1)}$ provides a conjugacy of period 2 while the sequence $\{\omega^{1/d} z, \omega z, \overline{\omega} z, \omega z, \overline{\omega} z, \dots\}$ provides a strictly preperiodic conjugacy provided we choose a branch of $z^{1/d}$ other than that which gives us $\overline{\omega}$ as a dth root of ω. Finally, if a is any irrational number and we let $\alpha = e^{2\pi i a}$, then the sequence $\{\alpha z, \alpha^d z, \alpha^{d^2} z, \alpha^{d^3} z, \dots\}$ gives a conjugacy where the sequence of mappings clearly never repeats. Following our scheme above, we shall call this kind of conjugacy an *aperiodic conjugacy.*

In view of what we have proved above, we see that the notion of conjugacy for random iteration will be strong enough for us to make many meaningful statements concerning similarities in the behaviour of different random dynamical systems. Our main result concerning conjugacy which we will make use of in constructing the counterexample we will outline in the next section is stated below.

Theorem 2.1. *Let $\{P_m\}_{m=1}^{\infty}$ be a bounded sequence of polynomials defined on all of $\overline{\mathbb{C}}$. Then $\{P_m\}_{m=1}^{\infty}$ is analytically conjugate to a bounded monic sequence of polynomials $\{\widetilde{P}_m\}_{m=1}^{\infty}$. Moreover, we may require that each $\widetilde{P}_m$ have a critical point at 0 for each m.*

Proof. Let K and M be the bounds respectively for the leading and the other coefficients of our sequence $\{P_m\}_{m=1}^{\infty}$ as given in the definition. The argument will proceed by constructing a 'partial conjugacy' $\varphi_0^n, \varphi_1^n, \dots, \varphi_n^n$ which satisfies $\varphi_m^n \circ P_m \circ (\varphi_{m-1}^n)^{-1} = \widetilde{P}_m^n$ for each $1 \leq m \leq n$ where $\widetilde{P}_m^n$ is monic and has a critical point at 0. This step can be thought of as a version of the 'pullback argument' which is a standard approach to constructing conjugacies in complex dynamics. We will then let $n \to \infty$ and apply a limiting argument to obtain the full result.

We begin by fixing $n \geq 0$, and setting $\varphi_n^n(z) = z$. Our induction assumption concerning the constants α_m^n is that $K^{-1/(d-1)} \leq |\alpha_m^n| \leq K^{1/(d-1)}$ for each $1 \leq m \leq n$ (here d is simply the degree of each of the polynomials in our sequence $\{P_m\}_{m=1}^{\infty}$). If $n = 0$, this condition is trivially satisfied and we have now of course finished constructing the partial conjugacy, otherwise we can assume from now on that $n \geq 1$. One can easily compute that in order for the leading coefficient of each of the polynomials $\widetilde{P}_m = \varphi_m^n \circ P_m \circ (\varphi_{m-1}^n)^{-1}$ to be 1, we must have that

$$|\alpha_{n-i}^n| = |a_{d,n}|^{1/d^i} |a_{d,n-1}|^{1/d^{i-1}} \cdots\cdots |a_{d,n-i+1}|^{1/d}, \qquad 0 \leq i < n$$

from which it follows immediately that $K^{\frac{-1}{d-1}} \leq |\alpha_{n-i}^n| \leq K^{\frac{1}{d-1}}$.

Turning now to the constant coefficients, it is again fairly easy to calculate that the requirement that the linear term of $\widetilde{P}_m$ be zero (which guarantees that there will be a critical point at 0) is equivalent to the condition that β_{n-i}^n, $1 \leq i \leq n$ satisfy the polynomial equation

$$\frac{d a_{d,n-i+1}(-\beta_{n-i}^n)^{n-1}}{\alpha_{n-i}^n} + \frac{(d-1)a_{d-1,n-i+1}(-\beta_{n-i}^n)^{n-2}}{\alpha_{n-i}^{n-1}} + \cdots\cdots + \frac{a_{1,n-i+1}}{\alpha_{n-i}} = 0.$$

If follows from the bounds on our sequence $\{P_n\}_{n=1}^{\infty}$ together with the bounds we have just established for the linear terms α_m^n, that if we use Rouché's theorem to compare the above polynomial with $\frac{(-1)^{n-1} d a_{d,n-i+1}}{\alpha_{n-i}^n} z^{n-1}$, then we can deduce that the constants β_m^n can be bounded uniformly in terms of the bounds K and M for $\{P_n\}_{n=1}^{\infty}$.

We therefore obtain a sequence of sequences of affine linear mappings $\{\varphi_m^n = \alpha_m^n z + \beta_m^n\}_{m=0}^{\infty}$ where the constants α_m^n and β_m^n are uniformly bounded. If we fix $m \geq 0$, we can find a subsequence n_k so that $\alpha_m^{n_k}$ and $\beta_m^{n_k}$ will converge. It then follows from the standard Cantor diagonalization procedure that we can find a subsequence n_k so that for every fixed m, $\alpha_m^{n_k}$ and $\beta_m^{n_k}$ will converge to some limits α_m and β_m respectively. If we now let $\varphi_m = \alpha_m z + \beta_m$, it follows that we can find a monic sequence of polynomials $\{\widetilde{P}_m\}_{m=1}^{\infty}$ each member of which has a critical point at 0 such that the sequence of functions $\{\varphi_m\}_{m=0}^{\infty}$ will provide the desired conjugacy between $\{P_m\}_{m=1}^{\infty}$ and $\{\widetilde{P}_m\}_{m=1}^{\infty}$ with the necessary properties.

From Theorem 2.1 we obtain the following corollary:

Corollary 2.1. *Let $\{P_n\}_{n=1}^{\infty}$ be a bounded sequence of quadratic polynomials. Then $\{P_n\}_{n=1}^{\infty}$ is conjugate to a sequence $\{\widetilde{P}_n\}_{n=1}^{\infty}$ of monic quadratic polynomials of the form $\widetilde{P}_n(z) = z^2 + c_n$ where $\{c_n\}_{n=1}^{\infty}$ is a sequence in l_∞. Also, for any conjugacy between two sequences of this type, the conjugating maps $\varphi_n = \alpha_n z + \beta_n$ must satisfy $|\alpha_n| = 1$, $\beta_n = 0$ for every n.*

Proof. The existence of the desired conjugacy follows immediately from Theorem 2.1.

For the uniqueness part, note that the linear coefficients α_n satisfy

$$\alpha_{n-1}^2 = \alpha_n, \qquad n \in \mathbb{N}$$

from which the result on the linear terms follows. For the constant terms, the fact that $\beta_n = 0$ for every n follows from the fact that each quadratic polynomial for the two sequences has a unique critical point at 0 together with the fact that the conjugacy will map critical points for the polynomials of one sequence to those of the other sequence.

This of course is the random analogue of the classical fact that any quadratic polynomial is conjugate to a polynomial of the form z^2+c and that no two distinct polynomials of this form can be conjugate to each other. There is therefore little loss of generality when working with quadratic polynomials, in restricting oneself to sequences of the form $P_n(z) = z^2 + c_n$ with $\{c_n\}_{n=1}^{\infty}$ a bounded sequence. For more on sequences of this type see [**1**, **2**, **3**] and [**6**].

Finally, we remark that using our earlier arguments, one may show that any two quadratic polynomials are conjugate in the random sense if and only if they are classically conjugate and if P^1 and P^2 are not classically conjugate to z^2, the mappings of any random conjugacy will all be a classical conjugacy except possibly at stage 0.

3. Proof of the main result.

We start by noting that it will suffice to construct a bounded polynomial sequence which has a Fatou component with the desired properties. It will then follow immediately from Theorem 2.1 and Proposition 2.1 that there is a monic polynomial sequence with a suitable Fatou component. The construction of the sequence which has the properties we require for Theorem 1.4 hinges mainly on the dynamics of the polynomial $P(z) = \lambda z(1-z)$ where $\lambda = e^{2\pi i\varphi}$ and $\varphi = (\sqrt{5}-1)/2$, the golden ratio. As is well-known, this polynomial gives rise to the famous golden mean Siegel disk. The origin is a neutral fixed point for P with multiplier λ and it lies in an invariant Fatou component U on which the dynamics are (classically) conjugate to a rotation by $2\pi\varphi$ which is of course an irrational multiple of 2π. It follows immediately from this we can find high iterates of P which come arbitrarily close to the identity on U (in the sense of uniform convergence on compact subsets of U). Our construction will rely heavily on this fact which we will use to control the distortion of a disk in U under the iterates of our sequence. The idea is to put the disk through ever longer cycles of shrinking and expanding where we first shrink it down to some arbitrarily small size and then expand it again to be approximately the same size and shape as it was originally. Our construction is by induction and we proceed to outline it below.

Induction Step: Stage 1. Let $D = D(0, r)$ be a closed disk of radius r and centre 0 for which $D(0, 2r)$ lies in U and let $\mu = r/2$. Note that this forces r to be less than $1/4$ as it is well-known that the critical point for P at $1/2$ lies on ∂U. We will construct our sequence of polynomials so that the interior of this disk will lie in a Fatou component for our sequence with the desired properties. Note that if we dilate D by a factor of $1/\mu = 2/r$, we obtain the disk $D(0, 2)$ whose boundary $C(0, 2)$, the circle centered at 0 of radius 2 lies entirely in the basin at infinity for P. On the other hand, if we dilate the disk $D(0, r^2/2)$ by the same factor, we simply obtain D which lies within U. The three polynomials we will use to construct our sequence are $P(z)$, $P(\mu z) = \lambda\mu z(1 - \mu z)$ and $P(z/\mu) = (\lambda/\mu)z(1 - z/\mu)$, the last two being P precomposed with dilations of μ and $1/\mu$ respectively. All sequences formed from these three polynomials are clearly bounded polynomial sequences in the sense of the definition given in Section 1, and we may therefore find an escape radius R so that if $|z| > R$, $Q_n(z) \to \infty$ as $n \to \infty$ where Q_n is any arbitrary composition of polynomials chosen from among these three. We will make use of these observations later when it comes to proving the second part of Theorem 1.4.

Our first induction step consists of first shrinking D by a factor of μ under the dilation μz, applying P $m_{1,1}$ times where $m_{1,1}$ is a natural number to be determined, then applying a dilation by a factor of $1/\mu$ and finally again applying P $m_{1,2}$ times where $m_{1,2}$ is also to be determined. We now define the first $m_{1,1} + m_{1,2}$ members of our sequence of polynomials by setting $P_1(z) = P(\mu z)$, $P_n(z) = P(z)$ for $1 < n \leq m_{1,1}$, $P_{m_{1,1}+1}(z) = P(z/\mu)$ and finally $P_n(z) = P(z)$ for $m_{1,1} + 1 < n \leq m_{1,1} + m_{1,2}$. Clearly it follows that $Q_{m_{1,1}}(z) = P^{\circ m_{1,2}} \circ z/\mu \circ P^{\circ m_{1,1}}(z) \circ \mu z$ and so this polynomial has the same effect on D as the process we outlined above.

The dilations clearly do not introduce any distortion on the image of D while by choosing $m_{1,1}$ and $m_{1,2}$ large enough, we may assume that $P^{\circ m_{1,1}}$ and $P^{\circ m_{1,2}}$ are as close to the identity on $D(0, 2r)$ as we wish. It follows from these two facts that we may make the image of D under $Q_{m_{1,1}}$ (i.e., $D_{m_{1,1}}$) as close to $D(0, r^2/2)$ and the image under $Q_{m_{1,1}+m_{1,2}}$ as close to D as we wish. To be more precise, we specify that if $\partial D = C(0, r)$, then we require that the Hausdorff distance between $C(0, r^2/2)$ and $Q_{m_{1,1}}(C(0, r))$ is $\leq r^2/8$ and that between $C(0, r)$ and $Q_{m_{1,1}+m_{1,2}}(C(0, r))$ is $\leq r(1/4 + 1/8) = 3r/8$. In addition, if we consider the circle $C(0, r)$ at time $m_{1,1}$ which certainly contains $Q_{m_{1,1}}(C(0, r))$, the image of this circle under the dilation z/μ is $C(0, 2)$ and so by making $m_{1,2}$ as large as we wish, we may ensure that the iterate of the circle under $Q_{m_{1,1}, m_{1,1}+m_{1,2}}$ lies entirely outside $D(0, R)$ and so is guaranteed to escape to infinity regardless of how we choose from among our three polynomials for the construction of the rest of the sequence. Finally, we let the number of polynomials in our sequence specified so far be N_1 so that in this case we simply have that $N_1 = m_{1,1} + m_{1,2}$.

Induction Hypothesis: Stage n. We suppose now that we have constructed the first N_n members of our sequence of polynomials. Each of these has again been chosen from among the three polynomials $P(z), P(\mu z)$ and $P(z/\mu)$. The origin has therefore remained fixed under our sequence so far and the effect of the compositions of our polynomials has been that of iterates of P interspersed with dilations by either μ or $1/\mu$. Our first assumption concerning these is that for each $1 \leq i \leq n$, Q_{N_{i-1},N_i} consists of a composition of $N_i - N_{i-1}$ polynomials (where in the case $i = 1$ we set $N_0 = 1$) which can be expressed as i dilations by μ each of which is followed by some iterate of P followed by i dilations by $1/\mu$ each of which is again followed by an iterate of P. Let us denote the number of these iterates of P by $m_{i,1}, m_{i,2}, \ldots, m_{i,i}, \ldots, m_{i,2i}$ and for $1 \leq j \leq 2i$, let us denote the sum of the first j of these numbers by $M_{i,j}$, i.e., we set $M_{i,j} = \sum_{k=1}^{j} m_{i,k}$. Our second assumption concerning the polynomials chosen so far is that all iterates of P chosen so far are close enough to the identity so that the Hausdorff distance between $Q_{N_{i-1}+M_{i,i}}(C(0,r))$ and $C(0,\mu^i r)$ is $\leq \frac{\mu^i r}{4}\sum_{j=0}^{2i-1} 2^{-j} = \mu^{i+1}\sum_{j=1}^{2i} 2^{-i}$ and that between $Q_{N_i}(C(0,r))$ and $C(0,r)$ is $\leq \frac{r}{4}\sum_{j=0}^{2i} 2^{-j}$. Finally, we make the assumption that at stage $N_{i-1}+M_{i,i}$, all points on the circle $C(0,\mu^{i-1}r)$ (which encloses a disk which contains $D_{N_{i-1}+M_{i,i}}$ in view of our assumptions above) are guaranteed to escape to infinity by ensuring that the image of this circle at time N_n lies entirely outside $D(0,R)$ where R is the escape radius from above.

Induction Step: Stage n+1. The $(n+1)$st step of the induction consists of defining the next members of our sequence to be $n+1$ steps each of which is a dilation by μ followed by a suitable iterate of P followed by a further $n+1$ steps each of which is a dilation by $1/\mu$ each of which is again followed by a suitable iterate of P. To be more precise, let $m_{n+1,1}, m_{n+1,2}, \ldots, m_{n+1,2n+2}$ be natural numbers to be determined and let $M_{n+1,i} = \sum_{j=1}^{i} m_{n+1,j}$ for each $1 \leq i \leq 2n+2$. Now let $N_{n+1} = N_n + M_{n+1,2n+2}$ and for $N_n + 1 \leq n \leq N_{n+1}$, define P_n by

$$P_n(z) = \begin{cases} P(\mu z) & n = N_n + 1 \\ P(\mu z) & n = N_n + M_{n+1,i} + 1, \quad 1 \leq i \leq n \\ P(z/\mu) & n = N_n + M_{n+1,i}, \quad n+1 \leq i \leq 2n+1 \\ P(z) & \text{otherwise.} \end{cases}$$

We can clearly choose the integers $m_{n+1,i}$ so that the corresponding iterates $P^{\circ m_{n+1,i}}$ are as close to the identity on $D(0,2r)$ as we wish. It then follows easily from the induction hypothesis and our initial assumption that $D(0, 2r) \subset U$ that we can ensure that the Hausdorff distance between $Q_{N_n+M_{n+1,n+1}}(C(0,r))$ and $C(0, \mu^{n+1}r)$ is $\le \mu^{n+2} \sum_{i=1}^{2n+2} 2^{-i}$ while that between $Q_{N_{n+1}}(C(0,r))$ and $C(0,r)$ is $\le \frac{r}{4} \sum_{i=0}^{2n+2} 2^{-i}$. Finally, for the same reasons, we may assume that if we consider the circle $C(0, \mu^n r)$ at time $N_n + M_{n+1,n+1}$ which certainly encloses a disk which contains

$$Q_{N_n+M_{n+1,n+1}}(C(0,r))$$

from above, if we now iterate with $Q_{N_n+M_{n+1,n+1},N_n+M_{n+1,2n+1}}$ we will obtain a curve which we can make as close to $C(0,r)$ in the Hausdorff topology as we desire. If we now dilate by a factor of $1/\mu$ and iterate with P the remaining $m_{n+1,2n+2}$ times and $m_{n+1,2n+2}$ is large enough, we can then ensure that the image of this circle at time $N_{n+1} = N_n + M_{2n+2}$ lies entirely outside $D(0, R)$ and so is guaranteed to escape to infinity regardless of how we construct our sequence in the future.

This completes the induction and the construction of our sequence. Our assumption concerning the distortion of D under iteration shows us that at the times N_n, the diameter of D_{N_n} is at most $r(1 + 1/4 \sum_{i=0}^{2n} 2^{-i}) < 3r/2$ which implies that the orbit of D must remain bounded under iteration and hence its interior must be contained in a Fatou component V for our sequence. On the other hand, one easily checks that $|Q'_{N_n}(0)| = 1$ from which it follows immediately that there is a nonconstant normal limit function on V. However, at the times $N_n + M_{n+1}$, $|Q'_{N_n+M_{n+1}}(0)| = \mu^{n+1}$ and so there must also be a constant normal limit function on V. Hence $\{P_n\}_{n=1}^{\infty}$ is a bounded polynomial sequence which has a Fatou component with the required properties for the first part of Theorem 1.4. For the second part of the result, we see from above that at the times N_n, one checks similarly to above that $\operatorname{diam} V_{N_n} \ge r$ for all $n \ge 0$ while the condition about the circles $C(0, \mu^n r)$ at times $N_n + M_{n+1,n+1}$ escaping to infinity ensures that $\operatorname{diam} V_{N_n+M_{n+1,n+1}} \le \mu^n r$. This shows that V has the required properties for the second part of the theorem. Finally, in view of our earlier remarks at the beginning of this section, it follows that there is also a monic sequence of polynomials which has a Fatou component with the desired properties for both parts of the theorem and with this the proof is complete.

Before finishing, we make a few remarks. The technique for constructing the desired monic sequence of polynomials is indirect, relying as it does on the notion of conjugacy. It appears that it is in fact rather difficult to proceed directly to construct such a sequence without using non-monic polynomials and appealing to conjugacy to obtain a monic sequence. One could for example try using two monic polynomials, one for the ‘contraction’ and the other for the ‘expansion’ parts of the cycles above which correspond to the steps in the induction. However, computer experiments show that attempts to use the dynamics of either hyperbolic or parabolic examples to do this do not seem to work. Roughly what happens is that although one can control the distortion of a Fatou component over one such cycle, the distortions one picks up from many of these cycles are not summable.

References

[1] R. Brück, *Geometric properties of Julia sets of the composition of polynomials of the form* $z^2 + c_n$, Pacific J. Math., **198** (2001), 347-371, MR 2002d:37078.

[2] ———, *Connectedness and stability of Julia sets of the composition of polynomials of the form* $z^2 + c_n$, J. London Math. Soc. (2), **61(2)** (2000), 462-470, MR 2001e:30040.

[3] R. Brück, M. Büger and S. Reitz, *Random iterations of polynomials of the form* $z^2 + c_n$*: Connectedness of Julia sets*, Ergodic Theory Dynam. Systems, **19** (1999), 1221-1231, MR 2000h:37064, Zbl 0942.37041.

[4] M. Büger, *Self-similarity of Julia sets of the composition of polynomials*, Ergodic Theory Dynam. Systems, **17** (1997), 1289-1297, MR 98k:5819, Zbl 0894.58029.

[5] L. Carleson and T.W. Gamelin, *Complex Dynamics*, Springer Verlag, Universitext: Tracts in Mathematics, 1993, MR 94h:30033, Zbl 0782.30022.

[6] M. Comerford, *Properties of Julia Sets Arising from Arbitrary Compositions of Monic Polynomials*, Ph.D. Thesis, Yale University, May 2001.

[7] ———, *Infinitely many grand orbits*, Michigan Math. J., **51** (2003), 47-57, CMP 1 960 920.

[8] A.E. Eremenko and M.Ju. Lyubich, *Examples of entire functions with pathological dynamics*, J. London Math. Soc., **36** (1987), 458-468, MR 89e:30047, Zbl 0601.30033.

[9] J.E. Fornaess and N. Sibony, *Random iterations of rational functions*, Ergodic Theory Dynam. Systems, **11** (1991), 687-708, MR 93c:58173, Zbl 0753.30019.

[10] S. Kolyada and L. Snoha, *Topological entropy of nonautonomous dynamical systems*, Random Comput. Dynam., **4** (1996), 205-233, MR 98f:58126, Zbl 0909.54012.

[11] J. Milnor, *Dynamics in One Complex Variable. Introductory Lectures*, Friedr. Vieweg & Sohn, Braunschweig, 1999. MR 2002i:37057, Zbl 0946.30013.

[12] N. Steinmetz, *Rational Iteration*, Walter de Gruyter, 1993, MR 94h:30035, Zbl 0773.58010.

[13] D. Sullivan, *Quasiconformal homeomorphisms and dynamics* I. *Solution of the Fatou-Julia problem on wandering domains*, Ann. of Math., **122** (1985), 401-418, MR 87i:58103, Zbl 0589.30022.

Received July 25, 2002 and revised October 30, 2002.

Department of Mathematics
University of California
202 Surge Bldg.
Riverside CA 92521-0135
E-mail address: marco@math.ucr.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

COMPUTATION OF SOME MODULI SPACES OF COVERS AND EXPLICIT $\mathcal{S}_n$ AND $\mathcal{A}_n$ REGULAR $\mathbb{Q}(T)$-EXTENSIONS WITH TOTALLY REAL FIBERS

Emmanuel Hallouin and Emmanuel Riboulet-Deyris

We study and compute an infinite family of Hurwitz spaces parameterizing covers of $\mathbb{P}^1_{\mathbb{C}}$ branched at four points and deduce explicit regular $\mathcal{S}_n$ and $\mathcal{A}_n$-extensions over $\mathbb{Q}(T)$ with totally real fibers.

Introduction.

In this paper, we study a family of covers of the projective line suggested to us by Gunter Malle, namely those covers of even degree $n \geq 6$, ramified over four points, with monodromy $\mathcal{S}_n$ and having branch cycle description $\mathbf{C} = (C_1, C_2, C_3, C_4)$ of type:

$$\left((n-2), 3, 2^{\frac{n-2}{2}}, 2^{\frac{n}{2}}\right).$$

Malle suspected the Hurwitz curves have genus zero for every $n \geq 6$ and some covers in the family have totally real fibers. A similar family was suggested and extensively studied by Dèbes and Fried in [**DF94**]. Unfortunately, their Hurwitz spaces happen to have a quadratic genus in n and only provide the expected regular extensions for degrees 5 and 7 (see [**DF94**], Theorem 4.11). Their work uses braid action formulae (see [**FV91**]) and complex conjugation action formulae (see [**FD90**], Proposition 2.3).

In this paper, we first follow the lines of Dèbes and Fried method and show that Malle's expectations were right. Then the second half of the paper is devoted to the explicit calculation of the concerned universal family of covers. To this end, we use an explicit version of Harbater's deformation techniques ([**Har80**]) as proposed in [**CG94**, **Cou99**]. As far as we know, it is the first time such an advanced method is used for computing an infinite family of covers. We present numerical results, obtained with *Magma*, showing the efficiency of the proposed method compared to the classical ones involving Buchberger algorithm. Indeed our computation reduces to solving *linear systems.*

In the first section, we recall results from the theory of Hurwitz spaces and (non)-rigidity methods developed in [**FV91**, **Völ96**, **Ful69**, **MM99**]. The second section is devoted to the combinatoric study of our family and arithmetical consequences of it using the method of Dèbes and Fried. In the third section, we show the existence of totally real $\mathcal{S}_n$ and $\mathcal{A}_n$ residual $\mathbb{Q}$-extensions in that family. Related to this totally real specialization, we find a very special point of the boundary of our Hurwitz space that shows very useful when computing an explicit model. This is done in the last section, by a deformation method.

We notice that a fallout of our construction is the existence of totally real $\mathbb{Q}$-extensions with Galois group $\mathcal{S}_n$ and $\mathcal{A}_n$ with four rational branched points (with only three branched points such extentions does not exist, cf. [**Ser92**] §8.4.3.). However an astute and effective construction can be found in [**Mes90**] to realize $\mathcal{S}_n$, with an odd integer $n \geq 3$, as the regular Galois group of a degree n extension of $\mathbb{Q}(T)$ having totally real fibers; this construction leads to a number of branch points linear in n.

We thank Jean-Marc Couveignes for numerous extremely helpful discussions about this work and the anonymous referee for very interesting suggestions and comments.

1. General framework.

Let us observe that the covers of our family have monodromy group primitive with a three cycle so it is $\mathcal{A}_n$ or $\mathcal{S}_n$ and it must clearly be the later. Therefore this family is parameterized by a coarse moduli space called Hurwitz space, denoted $\mathcal{H}'_4(\mathcal{S}_n, \mathbf{C})$. It is a quasi-projective regular (not a priori connected) variety over $\overline{\mathbb{Q}}$ with the following properties (see [**FV91**, **Völ96**, **Ful69**]):

- Since any conjugacy class of $\mathcal{S}_n$ is rational, $\mathcal{H}'_4(\mathcal{S}_n, \mathbf{C})$ is defined over $\mathbb{Q}$.
- Let F_4 be the configuration space of 4 points, e.g., $(\mathbb{P}^1_{\mathbb{C}})^4 \setminus \mathcal{D}$ where $\mathcal{D}$ denotes the discriminant variety. The map:

$$\begin{aligned} \phi:\ \mathcal{H}'_4(\mathcal{S}_n, \mathbf{C}) &\longrightarrow F_4(\mathbb{P}^1_{\mathbb{C}}) \\ h &\longmapsto (z_1, z_2, z_3, z_4) \end{aligned}$$

  where z_1, z_2, z_3, z_4 are the branched points (in the given order) of the cover corresponding to the point h, is a finite étale morphism defined over $\mathbb{Q}$.
- Since $\mathcal{S}_n$ has no center, the covers in our family have no automorphism so the moduli space $\mathcal{H}'_4(\mathcal{S}_n, \mathbf{C})$ is a fine one and for any $h \in \mathcal{H}'_4(\mathcal{S}_n, \mathbf{C})$ the associated cover p_h can be defined over $\mathbb{Q}(h)$ the field of definition of the point h.

As in §4.2 of [**DF94**], rather than looking at the full moduli space, we concentrate on a curve in it. Let us fix three points $z_1, z_2, z_3 \in \mathbb{P}^1(\mathbb{Q})$ and consider the curve $\mathcal{H}'_{(z_1,z_2,z_3)}$ obtained by the pullback:

$$\begin{tikzcd}
\mathcal{H}'_{(z_1,z_2,z_3)} \arrow[r] \arrow[d, "\varphi"'] & \mathcal{H}'_4(\mathcal{S}_n, \mathbf{C}) \arrow[d, "\phi"] \\
\mathbb{P}^1_{\mathbb{C}} \setminus \{z_1, z_2, z_3\} \arrow[r, "i"] & F_4(\mathbb{P}^1_{\mathbb{C}})
\end{tikzcd}$$

(the lower horizontal map i is $z \mapsto (z_1, z_2, z_3, z)$). If the three fixed points are rational, all the maps are defined over $\mathbb{Q}$ and the curve $\mathcal{H}'_{(z_1,z_2,z_3)}$ is also defined over $\mathbb{Q}$.

2. Combinatoric study of the Hurwitz curve.

In this section and the following, we will see that every element of our Nielsen class can be braid to another one $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ such that the product $\sigma_1\sigma_2$ is an n-cycle. In this case, the Tchebycheff polynomial T_n appears by coalescing σ_1 and σ_2. This property make the computation relatively painless (see §4.2). Then we gather the more information we can about the cover $\varphi : \mathcal{H}'_{(z_1,z_2,z_3)} \longrightarrow \mathbb{P}^1_{\mathbb{C}} \smallsetminus \{z_1, z_2, z_3\}$ from its combinatoric description. This way we prove that the genus of $\mathcal{H}'_{(z_1,z_2,z_3)}$ is zero and have many rational points.

2.1. Nielsen classes description. Let us fix the branched points $\underline{z} = (z_1, z_2, z_3, z_4) \in F_4(\mathbb{P}^1_{\mathbb{C}})$ and an homotopic base of $\mathbb{P}^1_{\mathbb{C}} \setminus \{z_1, z_2, z_3, z_4\}$. Using the topological classification of covers, elements of the fiber $\phi^{-1}(\underline{z})$ are in bijection with $\mathrm{sni}^{\mathrm{ab}}(\mathcal{S}_n, \mathbf{C})$ the strict absolute Nielsen class of type $(\mathcal{S}_n, \mathbf{C})$, that is:

$$\mathrm{sni}^{\mathrm{ab}}(\mathcal{S}_n, \mathbf{C}) = \Big\{(\sigma_1, \sigma_2, \sigma_3, \sigma_4) \in (\mathcal{S}_n)^4,\ \sigma_1\sigma_2\sigma_3\sigma_4 = 1,$$
$$\sigma_i \in C_i\ \forall i,\ \langle \sigma_1, \sigma_2, \sigma_3, \sigma_4 \rangle = \mathcal{S}_n \Big\} \Big/ \sim$$

where $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) \sim (\sigma'_1, \sigma'_2, \sigma'_3, \sigma'_4)$ means that there exists $\tau \in \mathcal{S}_n$ such that $\sigma'_i = \tau\sigma_i\tau^{-1}$ for all $1 \leq i \leq 4$.

We first enumerate the Nielsen class. To this end, we construct a representative system of this set of equivalence classes made of 4-uple $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ such that the product $\sigma_1\sigma_2$ is a cycle.

Since any two $(n-2)$-cycles are $\mathcal{S}_n$-conjugate, every element of $\mathrm{sni}^{\mathrm{ab}}(\mathcal{S}_n, \mathbf{C})$ has a representative with $\sigma_1 = (1 \dots n-2)$. Conjugating by a power of σ_1, we also assume that $\sigma_2 = (n-2\ \ k\ \ l)$. We now distinguish three cases.[1]

- *The case* $\{k, l\} = \{n-1, n\}$. Conjugating by a power of σ_1 and by $(n-1\ \ n)$, every element in that class has a unique representative with $\sigma_1 = (1 \dots n-2)$, and $\sigma_2 = (n-2\ \ n-1\ \ n)$. So $\sigma_1\sigma_2 = (1 \dots n)$

[1]Our convention for multiplying permutations is the following: $(12)(23) = (123)$.

and the enumeration reduces to finding all the permutations $(\sigma_4, \sigma_3) \in C_4 \times C_3$ such that $\sigma_4\sigma_3 = (1 \ldots, n)$ and $\langle \sigma_1, \sigma_2, \sigma_3, \sigma_4 \rangle = \mathcal{S}_n$. We need a lemma used several times further. It deals with relations in the diedral group D_m.

Lemma 2.1. *Let $m \leq n$ be even and let c be an m-cycle of $\mathcal{S}_n$. There is a bijection between nontrivial cycles of $c^{\frac{m}{2}}$ (i.e., sets of the form $\{x, c^{\frac{m}{2}}(x)\}$ where x belongs to the support of c), and the decompositions $c = \sigma\tau$ with σ a product of $\frac{m}{2}$ transpositions and τ a product of $\frac{m}{2} - 1$ transpositions. Therefore there are exactly $\frac{m}{2}$ such decompositions of c.*

Proof. Let x be an element of the support of c then one can verify that:

$$c = \underbrace{\prod_{i=1}^{\frac{m}{2}} \left(c^{1-i}(x)\, c^{i}(x)\right)}_{\sigma_{c,x}} \underbrace{\prod_{j=1}^{\frac{m}{2}-1} \left(c^{j}(x)\, c^{-j}(x)\right)}_{\tau_{c,x}}. \tag{1}$$

This is a decomposition associated to the set $\{x, c^{\frac{m}{2}}(x)\}$. Two such decompositions associated to x and y are equal if and only if $\{x, c^{\frac{m}{2}}(x)\} = \{y, c^{\frac{m}{2}}(y)\}$. On the other hand, let $c = \sigma\tau$ be a decomposition as in the lemma then one can show that it can be written as in (1) by considering an element x of the support of σ not belonging to the support of τ. $\square$

The group D_m is known to be the group of isometries of a regular polygon with m vertices. We can explain the previous relation geometrically: For each vertex x, the rotation (i.e., c) inducing an m-cycle on the vertices equals the composition of the unique reflection permuting the vertex x and its successor (i.e., $\sigma_{c,x}$) with the unique reflection fixing x (i.e., $\tau_{c,x}$).

Back to the enumeration of the Nielsen classes in the case where $\{k, l\} = \{n-1, n\}$, Lemma 2.1 shows that there are exactly $\frac{n}{2}$ such elements; this subset is denoted Class $\mathcal{A}$ in Table 1.

- *The case* $\#(\{k, l\} \cap \{n-1, n\}) = 1$. Conjugating by a power of σ_1 and by $(n-1\ \ n)$, every class has a unique representative with $k \in \{1, \ldots, n-3\}$, $\sigma_1 = (1 \ldots n-2)$ and $\sigma_2 = (k\ \ \ n-2\ \ \ n-1)$. Inspect $(1 \ldots k)(k+1 \ldots n-1) = \sigma_1\sigma_2 = \sigma_4\sigma_3$. The right side fixes n, and so $\sigma_3(n) = \sigma_4(n)$, and $\sigma_4\sigma_3$ also fixes $\sigma_4(n)$. So, $k = 1$ and $(1\ \ n)$ appear in both σ_3 and σ_4; the rests of the decompositions are given by Lemma 2.1 with $c = (2 \ldots n-1)$. These elements form the Class $\mathcal{B}$ in Table 1.
- *The case* $\{k, l\} \subset \{1, \ldots, n-3\}$. In that case $\sigma_1 = (1 \ldots n-2)$ and $\sigma_2 = (k\ \ l\ \ n-2)$, then:

$$\begin{cases} \sigma_1\sigma_2 = (1 \ldots k\ \ l+1 \ldots n-2\ \ k+1 \ldots l) & \text{if } k < l \\ \sigma_1\sigma_2 = (1 \ldots l)(l+1 \ldots k)(k+1 \ldots n-2) & \text{if } k > l. \end{cases}$$

Since $\sigma_1\sigma_2 = \sigma_4\sigma_3$ fixes $n-1$ and n, transitivity on $\{1,\dots,n\}$ implies $(n-1\ \ n)$ is not a disjoint cycle in both σ_3 and σ_4. Therefore $\sigma_4\sigma_3$ has $n-1, \sigma_4(n-1), n$ and $\sigma_4(n)$ as fixed points. This implies $k > l$ and two of the disjoint cycle lengths l, $k-1$ and $n-2-k$ are 1. Each of the three possibilities for σ_2 $(n-3\ \ 1\ \ n-2)$ $(2\ \ 1\ \ n-2)$ and $(n-3\ \ n-4\ \ n-2)$, are conjugate under $\langle\sigma_1\rangle$. With no loss, the Nielsen class representative has $\sigma_2 = (n-3\ \ n-4\ \ n-2)$. Then $\sigma_1\sigma_2 = (1\dots n-4)$ and in the support of σ_3 and σ_4, we find $(n\ \ n-2)(n-1\ \ n-3)$ or $(n\ \ n-3)(n-1\ \ n-2)$ which are conjugate under $(n-1\ \ n)$. Again, Lemma 2.1 for $c = (1\dots n-4)$ concludes. These elements form the Class $\mathcal{C}$ in Table 1.

The whole enumeration can be found in Table 1. Note that the three pointed classes have the following cardinalities:

$$\#\mathcal{A} = \frac{n}{2}, \qquad \#\mathcal{B} = \frac{n}{2} - 1 \qquad \text{and} \qquad \#\mathcal{C} = \frac{n}{2} - 2$$

therefore:

$$\#\,\mathrm{sni}^{\mathrm{ab}}(\mathcal{S}_n, \mathbf{C}) = 3\left(\frac{n}{2} - 1\right).$$

Concerning the Hurwitz cover, we have shown that:

Fact 2.2. The degree of the Hurwitz cover ϕ (or φ) equals $3\left(\frac{n}{2}-1\right)$.

Class $\mathcal{A}$	$a_i = \left[(1\dots n-2), (n-2\ \ n-1\ \ n), \tau_{(1\dots n),i}, \sigma_{(1\dots n),i}\right]$ with $1 \le i \le \frac{n}{2}$
Class $\mathcal{B}$	$b_i = \left[(1\dots n-2), (1\ \ n-2\ \ n-1), \nu\tau_{(2\dots n-1),i}, \nu\sigma_{(2\dots n-1),i}\right]$ with $\nu = (1\ \ n)$ and $2 \le i \le \frac{n}{2}$
Class $\mathcal{C}$	$c_i = \left[(1\dots n-2), (n-2\ \ n-3\ \ n-4), \nu\tau_{(1\dots n-4),i}, \nu\sigma_{(1\dots n-4),i}\right]$ with $\nu = (n\ \ n-2)(n-1\ \ n-3)$ and $1 \le i \le \frac{n}{2} - 2$

Table 1. The Nielsen classes (same notations as in Lemma 2.1).

2.2. Braiding action. In this paragraph, we compute the action of braids on the Nielsen class given in Table 1.

2.2.1. Generator of the braid group and braiding action. Let us denote by $\mathcal{H}_4(\mathcal{S}_n, \mathbf{C})$ the Hurwitz space parameterizing the same set of isomorphism classes of covers as $\mathcal{H}'_4(\mathcal{S}_n, \mathbf{C})$ but without ordering the branch points. This space can be endowed with a topology which is constructed in the same way as the one of $\mathcal{H}'_4(\mathcal{S}_n, \mathbf{C})$ (see [**FV91**] or [**Ful69**]). On the other hand, it maps onto $C_4(\mathbb{P}^1_{\mathbb{C}})$, the quotient of $F_4(\mathbb{P}^1_{\mathbb{C}})$ by the action of $\mathcal{S}_4$ on the coordinates:

$$\begin{tikzcd} \mathcal{H}'_4(\mathcal{S}_n,\mathbf{C}) \arrow[r] \arrow[d, "\phi"] & \mathcal{H}_4(\mathcal{S}_n,\mathbf{C}) \arrow[d, "\phi'"] \\ F_4(\mathbb{P}^1_{\mathbb{C}}) \arrow[r] & C_4(\mathbb{P}^1_{\mathbb{C}}). \end{tikzcd} \tag{2}$$

The fundamental group of $C_4(\mathbb{P}^1_{\mathbb{C}})$ is the Hurwitz braid group of index 4. It possesses a classical presentation (see [**Han89**] or [**Bir75**]):

$$\left\langle Q_1, Q_2, Q_3 \;\middle|\; \begin{array}{l} Q_1Q_3 = Q_3Q_1 \\ Q_1Q_2Q_1 = Q_2Q_1Q_2 \text{ and } Q_2Q_3Q_2 = Q_3Q_2Q_3 \\ Q_1Q_2Q_3^2Q_2Q_1 = 1 \text{ (sphere's relation)} \end{array} \right\rangle.$$

Denoting the fiber of ϕ' by $\mathrm{ni}^{\mathrm{ab}}(\mathcal{S}_n,\mathbf{C})$ the "unordered" Nielsen class associated to inertia's 4-uple $\mathbf{C}$, i.e., the quotient of $\mathrm{sni}^{\mathrm{ab}}(\mathcal{S}_n,\mathbf{C})$ by the action of $\mathcal{S}_4$ on the coordinates, we have:

Proposition 2.3 (Braiding action formula)**.** *For all $i = 1, 2, 3$, and for all $[\sigma_1,\sigma_2,\sigma_3,\sigma_4]$ in $\mathrm{ni}^{\mathrm{ab}}(\mathcal{S}_n,\mathbf{C})$, the monodromy (right) action of the braid Q_i for the previous generating system is given by the formula:*

$$[\sigma_1,\sigma_2,\sigma_3,\sigma_4].Q_i = [\dots,\sigma_i\sigma_{i+1}\sigma_i^{-1},\sigma_i,\dots]$$

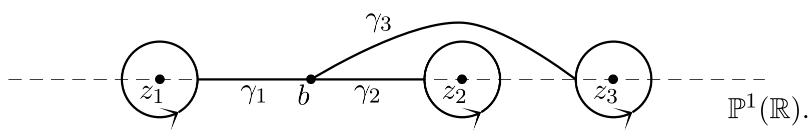
where $\sigma_i\sigma_{i+1}\sigma_i^{-1}$ is the i-th coordinate.

Generally speaking the diagram (2) permits us to express the monodromy of ϕ according to the one of ϕ'. This can be done explicitly because a presentation of the fundamental group of $F_4(\mathbb{P}^1_{\mathbb{C}})$ can be expressed in term of the one of $C_4(\mathbb{P}^1_{\mathbb{C}})$. We just give here the generators and refer to [**Han89**] for a complete system of relations:

$$t_{1,2} = {Q_1}^2, \qquad t_{2,3} = {Q_2}^2, \qquad t_{1,3} = Q_1{Q_2}^2{Q_1}^{-1}.$$

2.2.2. Monodromy action for the cover $\boldsymbol{\varphi}$. Let us recall that we just defined φ in Section 1 to be the pullback of ϕ by the monomorphism denoted by $i : \mathbb{P}^1_{\mathbb{C}} \smallsetminus \{z_1,z_2,z_3\} \hookrightarrow \mathcal{U}^4$. We now want to study the monodromy of this cover.

Let us choose $z_1 < z_2 < z_3$ on the real line $\mathbb{P}^1(\mathbb{R})$ and let us define the following "homotopic base" of $\mathbb{P}^1_{\mathbb{C}} \smallsetminus \{z_1,z_2,z_3\}$:



Then, as explained in details in Theorem 4.5 of [**DF94**], for an adapted generating system of braids[2] Q_1, Q_2, Q_3 on $C_4(\mathbb{P}^1_{\mathbb{C}})$, we can compute the morphism i_* induced by i on the respective homotopic groups in term of those two sets of generators:

$$i_*(\gamma_1) = Q_1^2 = t_{1,2}, \quad i_*(\gamma_2) = Q_2^2 = t_{2,3}, \quad i_*(\gamma_3) = Q_2^{-1}Q_3^2Q_2 = (t_{1,2}t_{2,3})^{-1}.$$

With these formulas, the computation of the monodromy of φ is deduced from the one of ϕ. We summarize in the next:

Proposition 2.4. *Using the notations of the Table* 1*, the monodromy action for the cover φ is:*

- *For the path γ_1:*

$$\left(a_{\frac{n}{2}}\ a_{\frac{n}{2}-1}\ \ldots\ a_1\right)\left(b_{\frac{n}{2}}\ b_{\frac{n}{2}-1}\ \ldots\ b_2\right)\left(c_{\frac{n}{2}-2}\ c_{\frac{n}{2}-3}\ \ldots\ c_1\right);$$

- *for the path γ_2:*

$$\left(a_{\frac{n}{2}-2}\ b_{\frac{n}{2}}\ b_{\frac{n}{2}-1}\ a_{\frac{n}{2}}\ c_{\frac{n}{2}-2}\right)\left(a_1\ c_{\frac{n}{2}-3}\right)\left(a_2\ c_{\frac{n}{2}-4}\right)\ldots\left(a_{\frac{n}{2}-3}\ c_1\right)$$

$$\left(a_{\frac{n}{2}-1}\right)\begin{cases}\left(b_2\ b_{\frac{n}{2}-2}\right)\left(b_3\ b_{\frac{n}{2}-3}\right)\ldots\left(b_{\frac{n}{4}-1}\ b_{\frac{n}{4}+1}\right)\left(b_{\frac{n}{4}}\right) & \textit{if } 4 \mid n,\\ \left(b_2\ b_{\frac{n}{2}-2}\right)\left(b_3\ b_{\frac{n}{2}-3}\right)\ldots\left(b_{\frac{n-2}{4}}\ b_{\frac{n+2}{4}}\right) & \textit{if } 4 \nmid n;\end{cases}$$

- *for the path $\gamma_1.\gamma_2$:*

$$\left(a_{\frac{n}{2}-1}\ a_{\frac{n}{2}-2}\ b_{\frac{n}{2}-1}\right)\left(c_1\ a_{\frac{n}{2}-4}\right)\left(c_2\ a_{(\frac{n}{2}-5)}\right)\ldots\left(c_{\frac{n}{2}-4}\ a_1\right)\left(c_{\frac{n}{2}-3}\ a_{\frac{n}{2}}\right)\left(c_{\frac{n}{2}-2}\ a_{\frac{n}{2}-3}\right)$$

$$\begin{cases}\left(b_{\frac{n}{2}}\ b_{\frac{n}{2}-2}\right)\left(b_{\frac{n}{2}-3}\ b_2\right)\left(b_{\frac{n}{2}-4}\ b_3\right)\ldots\left(b_{\frac{n}{4}+1}\ b_{\frac{n}{4}-2}\right)\left(b_{\frac{n}{4}}\ b_{\frac{n}{4}-1}\right) & \textit{if } 4 \mid n,\\ \left(b_{\frac{n}{2}}\ b_{\frac{n}{2}-2}\right)\left(b_{\frac{n}{2}-3}\ b_2\right)\ldots\left(b_{\frac{n+2}{4}}\ b_{\frac{n-2}{4}-1}\right)\left(b_{\frac{n-2}{4}}\right) & \textit{if } 4 \nmid n.\end{cases}$$

2.3. Ramification in the Hurwitz curve. In conclusion, the Hurwitz curve $\mathcal{H}'$ is a cover of $\mathbb{P}^1_{\mathbb{C}}$ of degree $3\left(\frac{n}{2}-1\right)$ ramified over three points, say $z_1, z_2, z_3 \in \mathbb{P}^1(\mathbb{Q})$ with ramification type described in Table 2.

Fact 2.5. The Hurwitz curve $\mathcal{H}'_{(z_1,z_2,z_3)}$ satisfies:

- It is irreducible;
- it is of genus zero and $\mathbb{Q}$-isomorphic to $\mathbb{P}^1_{\mathbb{Q}}$.

So, our family contains infinitely many covers defined over $\mathbb{Q}$.

Proof. The irreducibility comes from the transitivity of the braiding action. The Riemann-Hurwitz formula shows that the genus of $\mathcal{H}'_{(z_1,z_2,z_3)}$ is zero. We can also note that, for example, the point of ramification index 5 must be a rational one (this ramification index is isolated). Being defined over $\mathbb{Q}$, of

[2]In our situation the 3 points z_1, z_2 and z_3 are on the real line, so the choice of Q_1, Q_2, Q_3 is the standard one (see [**Han89**] for example). In general the 3 generators Q_i must be precisely given. We point out that this choice just depends on a given path through z_1, z_2, z_3 in this order.

if $4 \mid n$		
$\frac{n}{2}$ $\frac{n}{2}-1$ $\frac{n}{2}-2$	5 1^2 $2^{\frac{3n}{4}-5}$	3 $2^{\frac{3n}{4}-3}$
z_1	z_2	z_3
if $4 \nmid n$		
$\frac{n}{2}$ $\frac{n}{2}-1$ $\frac{n}{2}-2$	5 1 $2^{\frac{3(n-6)}{4}}$	3 1 $2^{\frac{3n-14}{4}}$
z_1	z_2	z_3

Table 2. Ramification types over z_1, z_2, z_3 in the Hurwitz curve (double line stands for repeated ramification points of same index).

genus zero and with a rational point, $\mathcal{H}'_{(z_1,z_2,z_3)}$ is necessarily $\mathbb{Q}$-isomorphic to $\mathbb{P}^1_{\mathbb{Q}}$. In particular, there are covers in our family defined over $\mathbb{Q}$. $\square$

3. Existence of totally real $\mathcal{S}_n$ and $\mathcal{A}_n$-extensions.

There is still a question left: Does our family contain elements defined over $\mathbb{Q}$ with totally real fibers?

In this section, we use complex conjugation action on fibers as describe by Dèbes and Fried (see Theorem 2.4 of [**FD90**] or Proposition 2.3 of [**DF94**]) and prove adapted formulae to our family and to our choice of homotopic basis.

3.1. Covers in the family with totally real fiber. We consider a finite algebraic cover $p : \mathcal{C} \to \mathbb{P}^1_{\mathbb{C}}$ ramified over four ordered real points $z_2 < z_3 < z_4 < z_1 \in \mathbb{P}^1(\mathbb{R})$, we fix $z_0 \in \mathbb{P}^1(\mathbb{R})$ a real base point between z_3 and z_4 and we denote by F the fiber $p^{-1}(z_0)$.

The complex conjugation will play a crucial role; let denote this conjugation by a bar; for example $\bar{p} : \bar{\mathcal{C}} \to \mathbb{P}^1_{\mathbb{C}}$ is the cover obtained from the first one by complex conjugation and $\overline{F}$ is its fiber above z_0. Let $c : F \to \overline{F}$ be the bijection induced by the complex conjugation.

- The complex conjugation also acts on the topological fundamental group by left composition (thank you complex conjugation for being continuous!). The fundamental group $\pi_1(\mathbb{P}^1_{\mathbb{C}} \setminus \{z_1, z_2, z_3, z_4\}, z_0)$ is simply denoted by π_1. In the rest of this paper, when we refer to the standard homotopic basis of π_1, we always mean the one drawn in the Figure 1.

  So we have $\gamma_1\gamma_4\gamma_3\gamma_2 = 1$ and the complex conjugation acts as follows:

$$(3) \qquad \overline{\gamma_1} = \gamma_4^{-1}\gamma_1^{-1}\gamma_4 \qquad \overline{\gamma_4} = \gamma_4^{-1} \qquad \overline{\gamma_3} = \gamma_3^{-1} \qquad \overline{\gamma_2} = \gamma_3\gamma_2^{-1}\gamma_3^{-1}.$$

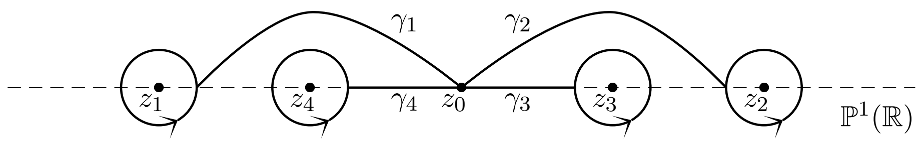


Figure 1. The standard homotopic basis.

- Since complex conjugation is a continuous morphism of $\mathbb{C}$, the monodromy of the cover $\bar{p}$ can be deduced from the monodromy of p. More precisely, if we denote by $\rho : \pi_1 \to \mathcal{S}_F$ and $\bar{\rho} : \pi_1 \to \mathcal{S}_{\overline{F}}$ the two monodromy (anti)morphisms, then we have:

$$\forall \gamma \in \pi_1, \qquad \bar{\rho}(\gamma) \circ c = c \circ \rho(\bar{\gamma}). \tag{4}$$

- From the Weil descent criterion (which boils down to the use of Artin theory because the extension $\mathbb{C}/\mathbb{R}$ is galois and finite) we know that the cover p can be defined over $\mathbb{R}$ if and only if there exists an isomorphism Ω such that:

$$\begin{tikzcd} \mathcal{C} \arrow[rr, "\Omega"] \arrow[dr, "p"'] & & \bar{\mathcal{C}} \arrow[dl, "\bar{p}"] \\ & \mathbb{P}^1_{\mathbb{C}} & \end{tikzcd} \qquad \text{and} \qquad \overline{\Omega} \circ \Omega = \mathrm{Id}$$

(the last condition is a cocycle condition).

From these three points, we can deduce a completely combinatoric criterion for the descent to $\mathbb{R}$ and for the existence of totally real fibers:

Theorem 3.1. *Let $p : \mathcal{C} \to \mathbb{P}^1_{\mathbb{C}}$, z_0, and F be as above. Let $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) \in \mathcal{S}_n^4$ be the branch cycle description of p — i.e., $\sigma_i = \rho(\gamma_i)$ where $\rho : \pi_1 \to \mathcal{S}_F$ is the monodromy morphism — and $G = \langle \sigma_1, \sigma_2, \sigma_3, \sigma_4 \rangle \subset \mathcal{S}_F$ the monodromy group of p.*

1. *The cover p is defined over $\mathbb{R}$ if and only if there exists an involution $\tau \in \mathcal{S}_F$ such that:*

$$\sigma_4 \sigma_1^{-1} \sigma_4^{-1} = {}^{\tau}\sigma_1, \qquad \sigma_4^{-1} = {}^{\tau}\sigma_4, \qquad \sigma_3^{-1} = {}^{\tau}\sigma_3, \qquad \sigma_3^{-1}\sigma_2^{-1}\sigma_3 = {}^{\tau}\sigma_2.$$

2. *If that is the case and if moreover the cover p has no automorphism — i.e., if the centralizator of G in $\mathcal{S}_F$ is trivial —, then the fiber F is totally real if and only if $\tau = \mathrm{Id}$.*

Proof. Corresponding to the isomorphism Ω of the Weil descent criterion there is a bijection $\omega : F \to \overline{F}$; in term of ω the conditions of the criterion are:

$$[\forall \gamma \in \pi_1, \quad \omega \circ \rho(\gamma) = \bar{\rho}(\gamma) \circ \omega] \quad \text{and} \quad \left[(c^{-1} \circ \omega)^2 = \mathrm{Id}\right]$$

(the last condition is just the cocycle one). Let $\tau = c^{-1} \circ \omega$. This is an involution of $\mathcal{S}_F$ which by (4) and (3) satisfies the expected relations on the σ_i (be careful: The monodromy is an anti-morphism) if and only if p is defined over $\mathbb{R}$.

Secondly, assuming that p is defined over $\mathbb{R}$, then the isomorphism Ω is an automorphism; so if moreover p has no automorphism, Ω, and therefore ω, must be identity. Furthermore, the conjugation c can now be viewed as a permutation of F. In conclusion, the bijection τ introduced in 1 satisfies $\tau = c^{-1}$ and the fiber F is totally real if and only if $\tau = \mathrm{Id}$. □

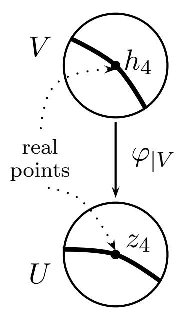
Having this result in mind, we can now go back to our family. Recall that the covers of our family have no automorphism (since the center of $\mathcal{S}_n$ is trivial). It is not difficult to verify that the Nielsen class:

$$a_{\frac{n}{2}-1} = \begin{cases} \sigma_1 = (12 \dots n-2) \\ \sigma_2 = (n-2 \;\; n-1 \;\; n) \\ \sigma_3 = \left[\prod_{i=1}^{\frac{n}{2}-2} (i \;\; n-i-2)\right] (n-2 \;\; n)(\frac{n}{2}-1)(n-1) \\ \sigma_4 = \left[\prod_{i=1}^{\frac{n}{2}-1} (i \;\; n-i-1)\right] (n-1 \;\; n) \end{cases}$$

satisfies the previous theorem and is the only one in that case. This calculation is very easy and is surely what motivated G. Malle to suggest to us this example.

Fact 3.2. Our family contains covers defined over $\mathbb{R}$ with an interval of non ramified points with totally real fibers.

3.2. Totally real $\mathcal{S}_n$-extensions. At this point we know that our family contains some covers defined over $\mathbb{Q}$ and some others with totally real fibers; we want to prove that there are covers satisfying both properties. We will have to move one of the ramification points.



Suppose that three of the ramification points $z_1, z_2, z_3 \in \mathbb{P}^1_{\mathbb{C}}(\mathbb{Q})$ are fixed and let the fourth one z_4 move on $\mathbb{P}^1(\mathbb{R})$ between z_1 and z_3. Thanks to the previous section, for all such z_4, there is a unique point $h_4 \in \varphi^{-1}(z_4)$ which represents a cover having totally real fibers above $]z_2, z_3[$ (namely the cover with branch cycle description $a_{\frac{n}{2}-1}$ with respect to a standard homotopic basis). We choose U and V two neighborhoods of z_4 and h_4 respectively such that $\varphi_{|V}$ becomes an homeomorphism from V to U. For every $z \in U \cap \mathbb{P}^1(\mathbb{R})$, the covering corresponding to $\varphi_{|V}^{-1}(z)$ satisfies the preceding descent criteria; so it is defined over $\mathbb{R}$ and has a real interval of specialization. Thus we have $\varphi_{|V}^{-1}(U \cap \mathbb{P}^1(\mathbb{R})) \subset V \cap \mathbb{P}^1(\mathbb{R})$. But rational points are dense in $\mathbb{P}^1(\mathbb{R})$ so we can find rational points in $V \cap \mathbb{P}^1(\mathbb{R})$; all the corresponding covers are defined over $\mathbb{Q}$ with a complete segment of totally real fibers:

Fact 3.3. Our family contains a rational nonempty interval of covers defined over $\mathbb{Q}$ each having an interval of totally real fibers.

Let $p : \mathbb{P}^1_{\mathbb{Q}} \to \mathbb{P}^1_{\mathbb{Q}}$ be such a cover. Since an interval is not a thin set, by Hilbert irreducible theorem, one can find irreducible and totally real specializations.

Proposition 3.4. *There exists totally real $\mathcal{S}_n$-extensions of $\mathbb{Q}$ of degree n.*

3.3. Totally real $\mathcal{A}_n$-extensions. From the previous construction, we now want to deduce the same kind of result for the group $\mathcal{A}_n$. Let $\mathcal{C} \to \mathbb{P}^1_{\mathbb{Q}}$ be one of our covers over $\mathbb{Q}$ with totally real fibers and $\mathcal{C}^{\mathrm{gal}} \to \mathbb{P}^1_{\mathbb{Q}}$ its Galois closure. The Galois group is still $\mathcal{S}_n$ because the arithmetic monodromy group contains the geometric one which is $\mathcal{S}_n$ too. We consider $\mathcal{D} = \mathcal{C}^{\mathrm{gal}}/\mathcal{A}_n$ as in the following diagram:

$$\begin{tikzcd}
\mathcal{C}^{\mathrm{gal}} \arrow[d, "\mathcal{A}_n"'] \arrow[dr] & \\
\mathcal{D} \arrow[d, "2"'] & \mathcal{C} \arrow[dl, "n"] \\
\mathbb{P}^1. &
\end{tikzcd}$$

Since two of the inertia permutations are even and the two others are odd, there are only two branched points in the cover $\mathcal{D} \to \mathbb{P}^1$, i.e., z_1 and z_3 or z_4 according to $n \equiv 0$ or $2 \pmod 4$. Therefore, by Riemann-Hurwitz formula, $\mathcal{D}$ is a genus zero curve; since it has at least one rational point (for example, one of the two ramified points), it is also $\mathbb{Q}$-isomorphic to $\mathbb{P}^1_{\mathbb{Q}}$ (this kind of argument looks like the so-called "double group trick", see [**Ser92**]).

This gives $\mathcal{A}_n$ with totally real fibers. Moreover the conjugacy classes for $\mathcal{C}^{\mathrm{gal}} \to \mathcal{D}$ are of the type:

$$\left(\left(\frac{n}{2}-1\right)^2, 3, 3, 2^{\frac{n-2}{2}}, 2^{\frac{n-2}{2}}\right) \qquad \text{or}$$
$$\left(\left(\frac{n}{2}-1\right)^2, 3, 3, 2^{\frac{n}{2}}, 2^{\frac{n}{2}}\right)$$

according to the parity of $\frac{n}{2}$. In conclusion:

Proposition 3.5. *There exist totally real $\mathcal{A}_n$-extensions of $\mathbb{Q}$ of degree n.*

At the end of this paper, we give an explicit version of both Propositions 3.4 and 3.5.

4. Explicit computation.

In this section, we compute the Hurwitz space and the universal curve for our family of covers. We fix once and for all three rational points $z_1 < z_3 < z_2$. To ease notations we denote by $\mathcal{H}$ the curve $\mathcal{H}'_{(z_1,z_2,z_3)}$.

4.1. The universal curve and a choice of coordinates. Because the covers in our family have no automorphism, by [**FV91**] or [**BF83**], there exists a universal curve $\mathcal{S}$ and a fibration:

$$\begin{tikzcd} \mathcal{H} \arrow[d, "\varphi"] & & \mathcal{S} \arrow[ll, "\Pi"'] \arrow[d, "\Phi"] \\ \mathbb{P}^1_{\mathbb{C}} \smallsetminus \{z_1, z_2, z_3\} & & F_2(\mathbb{P}^1_{\mathbb{C}} \smallsetminus \{z_1, z_2, z_3\}) \arrow[ll, "\pi"'] \end{tikzcd}$$

where:

- $\mathcal{S}$ is a smooth quasi-projective surface over $\mathbb{Q}$,
- $F_2(\mathbb{P}^1_{\mathbb{C}} \smallsetminus \{z_1, z_2, z_3\})$ denotes the quasi-projective variety defined by the ordered pairs $(u, v) \in \mathbb{P}^1_{\mathbb{C}} \smallsetminus \{z_1, z_2, z_3\}$ with $u \neq v$,
- π is the morphism obtained by forgetting the second coordinate v,
- the vertical arrows are finite and étale morphisms of varieties, all defined over $\mathbb{Q}$.

The morphism π admits a well-known projective completion that is still denoted by π. This is the canonical morphism from the (projective) moduli space of curves of genus zero with 5 marked points $\mathcal{M}_{0,5}$ to the one with 4 marked points $\mathcal{M}_{0,4}$ (we refer to [**GHP88**] for a highly comprehensive study of the spaces $\mathcal{M}_{0,n}$ from an algebraic view point. A lot of ideas contained in this paper are used here). We define $\overline{\mathcal{S}}$ to be the normalization of $\mathcal{M}_{0,5}$ in $\mathbb{Q}(\mathcal{S})$. We obtain this way a commutative diagram between smooth projective varieties defined over $\mathbb{Q}$ which extends the previous one:

$$\begin{tikzcd} \overline{\mathcal{H}} \arrow[d, "\varphi"] & \overline{\mathcal{S}} \arrow[l, "\Pi"'] \arrow[d, "\Phi"'] \\ \mathcal{M}_{0,4} & \mathcal{M}_{0,5}. \arrow[l, "\pi"'] \end{tikzcd}$$

In order to choose a system of coordinates on the universal curve, we need to define another Hurwitz curve, a little bigger than $\mathcal{H}$, namely the moduli space of covers in our family with a marking of an unramified point above z_3. This space denoted by $\mathcal{H}^{\bullet}$ is a degree 2 cover of $\mathcal{H}$ by the forgetting map. We check that this cover is connected. There is a universal curve $\mathcal{S}^\bullet \to \mathcal{H}^\bullet$ obtained by base extension of $\mathcal{S} \to \mathcal{H}$. A normalization, as in the previous paragraph, provides the following commutative diagram of projective varieties over $\mathbb{Q}$:

$$\begin{tikzcd} \overline{\mathcal{H}^\bullet} \arrow[d, "\varphi"] & \overline{\mathcal{S}^\bullet} \arrow[l, "\Pi^\bullet"'] \arrow[d, "\Phi"'] \\ \mathcal{M}_{0,4} & \mathcal{M}_{0,5}. \arrow[l, "\pi"'] \end{tikzcd}$$

Now, we want to choose adapted coordinates on those spaces. To begin with, let's recall that there exist four sections s_1, s_2, s_3 and s_4 of π, corresponding to the four marked points (see [**GHP88**], Section 3). We define x and y to be the coordinates on $\mathcal{M}_{0,5}$ such that $x = y = \infty$ on s_1, $x = 1$ on s_2, $x = y = 0$ on s_4 and $y = 1$ on s_3. Then we set $\lambda := \frac{x}{y} \in \mathbb{C}(\mathcal{M}_{0,4})$.

We also need coordinates on the universal curve $\overline{\mathcal{S}^\bullet}$. The fibration $\overline{\mathcal{S}^\bullet} \to \overline{\mathcal{H}^\bullet}$ admits four sections corresponding to the four points A, B, C, D (see Table 3). This amounts to saying that these four points are defined over $\mathbb{C}(\overline{\mathcal{H}^\bullet})$. This is clear for A and B because they are isolated. This is also true for C and D by definition of $\overline{\mathcal{S}^\bullet}$. We define the function X to be the unique coordinate taking values $\infty, 1, 0$ at A, B, C respectively. We define Y to be the unique coordinate taking values $\infty, 0, 1$ at A, C, D respectively. We set $\mu := \frac{X}{Y} \in \mathbb{C}(\overline{\mathcal{H}^\bullet})$. This situation is summarized in Table 3.

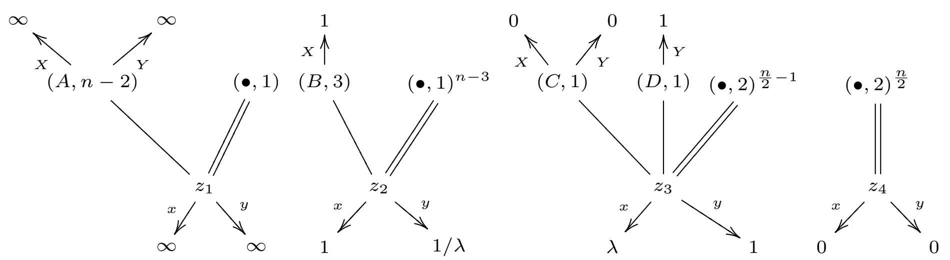


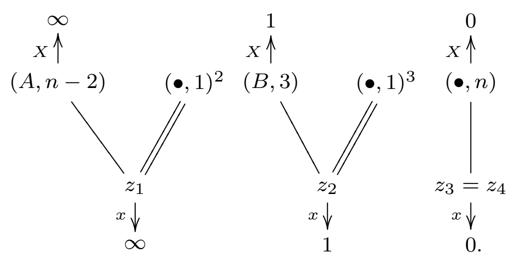
Table 3. Pointing the covers of $\mathcal{H}$ and choice of coordinates.

In order to compute an algebraic model for the cover Φ, we first exhibit an explicit model for a degenerate cover. In other words we compute the residual morphism induced by Φ on the fiber over a boundary point of $\mathcal{M}_{0,4}$ (that is corresponding to z_1, z_2 or z_3). Next, we can rebuilt the entire family by formal deformation of this residual morphism viewed as a morphism of curves on a formal power series ring.

4.2. Degenerate covers and their computation. Let b be a boundary point of $\mathcal{M}_{0,4}$. We fix a point, e.g., $b = z_3$. Since the key point of this section is the explicit algebraic structure of the compactification of moduli spaces of curves of genus zero, we will assume that the reader has some familiarities with these notions. We just recall (see [**GHP88**], Section 3) that the fiber $\mathcal{C}_b = \pi^{-1}(b)$ is a stable 4-pointed tree of projective lines made of two irreducible components. Let us denote by $\mathcal{C}_{b,1}$ the irreducible component of $\mathcal{C}_b$ containing the two closed points z_1 and z_2, and $\mathcal{C}_{b,2}$ the other one. For each point $h \in \overline{\mathcal{H}^\bullet}$ such that $\varphi(h) = b$, set $\mathcal{D}_h = (\Pi^\bullet)^{-1}(h)$. Then the restriction Φ_h of Φ on $\mathcal{D}_h$ is a cover of nodal curves (see for example Figure 2). We denote by $\Phi_{h,i}$ the restriction Φ_h to $\mathcal{D}_{h,i} = \Phi_h^{-1}(\mathcal{C}_i) \cap \mathcal{D}_h$.

For $i = 1, 2$ the morphism $\Phi_{h,i}$ is finite and the ramification locus is contained in the union of the two marked points and the singular point. Now the monodromy can classically be deduced (see [**Cou00**] for example) from the one of the nondegenerate covers in a small neighborhood of h. More precisely, let V be a small enough neighborhood of h (for the complex topology). If the branch cycle description of covers parameterized by $V \setminus \{h\}$ is given by a 4-uple $[\sigma_1, \sigma_2, \sigma_3, \sigma_4]$ then the branch cycle description of $\Phi_{h,1}$ (respectively $\Phi_{h,2}$) is $[\sigma_1\sigma_2, \sigma_3, \sigma_4]$ (respectively $[\sigma_1, \sigma_2, \sigma_3\sigma_4]$).

Let us now concentrate on the computation of an algebraic model for the covers $\Phi_{h,1}$ and $\Phi_{h,2}$. Recall, from the beginning of this section, that we have fixed three rational points $z_1 < z_3 < z_2$. For any $z_4 \in]z_1, z_3[$ let p_{z_4} be the cover of $\mathbb{P}^1 \setminus \{z_1, z_4, z_3, z_2\}$ with monodromy $a_{\frac{n}{2}-1}$ in the homotopic basis represented in Figure 1. Letting z_4 tends to z_3 we define a point h in the boundary of $\overline{\mathcal{H}^\bullet}$. For this h, the ramification data for $\Phi_{h,2}$ is:



This cover is a Padé approximant and for the couple of coordinates chosen, we have:

$$x = \Phi_{h,2}(X) = \frac{X^n}{\frac{n(n-1)}{2}\left(X^2 - 2\frac{n-2}{n-1}X + \frac{n-2}{n}\right)}. \tag{5}$$

Similarly for $\Phi_{h,1}$, we find:

$$y = \Phi_{h,1}(Y) = \frac{1}{2}T_n(2Y-1) + \frac{1}{2}, \tag{6}$$

where T_n denotes the Tchebycheff polynomial of order n.

Therefore a very simple model for this degenerate cover is known for every even n. We stress this decisive fact for the relevance of our approach.

4.3. Effective deformation and formal patching. In order to built the entire family, we now formally deform the previous degenerate cover. The deformation technique for covers appeared in [**Ful69**] and were then developed by D. Harbater for the study of the inverse Galois problem over complete local fields (see, e.g., [**Har80**, **Har87**]). In [**Wew99**], S. Wewers gives a presentation of the technique of deformation very well adapted to our purpose.

4.3.1. Computation of a formal model using effective deformation. Let us denote by R the complete local ring of $\overline{\mathcal{H}^{\bullet}}$ at the point h, namely $\mathbb{C}[[\mu]]$ because μ is a local parameter at h. By base extension $\mathcal{S}_R = \overline{\mathcal{S}^{\bullet}} \times_{\overline{\mathcal{H}^{\bullet}}} \mathrm{Spec}(R)$ and $\mathcal{C}_R = \mathcal{M}_{0,5} \times_{\mathcal{M}_{0,4}} \mathrm{Spec}(R)$ are *projective nodal curves* whose special fibers are nothing else than $\mathcal{D}_h$ and $\mathcal{C}_b$ respectively. The cover Φ induces a *tame admissible cover* from $\mathcal{S}_R$ to $\mathcal{C}_R$ which is a deformation of the previous degenerate cover Φ_h represented in Figure 2.

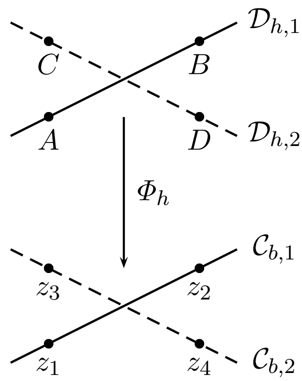


Figure 2. The degenerate cover Φ_h.

We describe the *deformation datum* associated to our deformation as it is explained in the paragraph of [**Wew99**] entitled *The main result* (pages 240-241). Using the preceding and putting $z = 1/y$, the curve $\mathcal{C}_R$ has affine equation $xz = \lambda$. So there is only one ordinary double point δ, i.e., $(0:0:1)$. Moreover, the complete local ring $\mathcal{O}_{\mathcal{C}_R,\delta} \simeq R[[x,z]]/\langle xz - \lambda\rangle$. We also have a *mark* on $\mathcal{C}_R$, namely the horizontal divisor defined by the four generic branched points

$$(0:1:0), \quad (1:\lambda:1), \quad (1:0:0), \quad (\lambda:1:1)$$

and this divisor is étale over $\mathrm{Spec}(R)$. The curve $\mathcal{S}_R$ has also a unique singular point Δ whose local ring is isomorphic to $R[[X,Z]]/\langle XZ - \mu\rangle$ where $Z = 1/Y$. Moreover, due to the ramification of the degenerate cover, we know that $X^n \sim x$, $Z^n \sim z$ and $\mu^n \sim \lambda$ (where $\sim$ means equal up to a unit factor). Therefore the *deformation datum* consists only in $\mu = XZ \in R$.

In concrete terms, effective patching permits us to compute a model of our family over a Puiseux series field $\mathbb{C}((\mu))$ where this field is nothing else than the completion of $\mathbb{C}(\overline{\mathcal{H}^\bullet})$ at the point h. We will call this model an formal one.

In view of our choice of coordinates, the model we are looking for has the following form:

$$\begin{aligned}
S(X) &= \frac{(X^{\frac{n}{2}} + \alpha_{\frac{n}{2}-1}X^{\frac{n}{2}-1} + \cdots + \alpha_0)^2}{\gamma(X^2+\beta_1 X+\beta_0)} = \frac{S_0(X)}{S_\infty(X)} \\
S(X)-1 &= \frac{(X-1)^3(X^{n-3}+\delta_{n-4}X^{n-4}+\cdots+\delta_0)}{\gamma(X^2+\beta_1 X+\beta_0)} = \frac{S_1(X)}{S_\infty(X)} \\
S(X)-\lambda &= \frac{X(X-\varepsilon_0)(X^{\frac{n}{2}-1}+\eta_{\frac{n}{2}-2}X^{\frac{n}{2}-2}+\cdots+\eta_0)^2}{\gamma(X^2+\beta_1 X+\beta_0)} = \frac{S_\lambda(X)}{S_\infty(X)}
\end{aligned}$$

where the $2n$ coefficients $\alpha_i, \beta_j, \delta_k, \eta_l$, ε_0 and γ have to be found in $\mathbb{C}(\overline{\mathcal{H}^\bullet})$. From the three equalities above, one can deduce that $S_0(X) - S_1(X) - S_\infty(X) = 0$ and that $S_0(X) - S_\lambda(X) - \lambda S_\infty(X) = 0$. This gives us a system of $2n$ polynomials in $2n$ variables and with coefficients in $\mathbb{C}[\mu]$ (because λ is a polynomial in μ) which should be satisfied by our $2n$ coefficients.

The knowledge of the degenerate cover gives us first order μ-adic development for all the coefficients. Then computing higher orders is just a careful application of the Newton-Hensel lemma.

4.3.2. Computation of an algebraic model from the formal one. First of all, we change the coordinate X by an homography which fixes 1 and ∞ so as to cancel the coefficient $\alpha_{\frac{n}{2}-1}$ in the polynomial $S_0(X)$:

$$X \longleftarrow \left(1 + \frac{\alpha_{\frac{n}{2}-1}}{\frac{n}{2}}\right) X - \frac{\alpha_{\frac{n}{2}-1}}{\frac{n}{2}}.$$

This natural normalization turns all the coefficients in our model to be defined over $\mathbb{C}(\mathcal{H})$. Hopefully, we have noticed the all the new coefficients are now power series in μ^2; so we are well back on $\mathcal{H}$!

Then the last part of the computation consists in deriving an algebraic model over $\mathbb{C}(\mathcal{H})$ from the preceding formal one defined over $\mathbb{C}((\mu^2))$ (which, we recall; is the completion of $\mathbb{C}(\mathcal{H})$ with respect to a point of $\mathcal{H}$). Theoretically speaking, this steps is based on the Artin's algebraization theorem. From a computational point of view, using the known μ-adic approximations of the coefficients, we should:

- First deduce a generator of $\mathbb{C}(\mathcal{H})$;
- and then express all the coefficients as rational fractions in this generator.

Even if the first step can be done systematically (as is explained in [**Cou99**]), we just guess a generator T among the coefficients and compute all the other coefficients in function of T. If we know the μ-adic approximations of T and of every coefficients C with enough accuracy, finding such expression is just a matter of linear algebra; indeed, for increasing values of the degree d, we have to solve in $\alpha_i, \beta_j \in \mathbb{C}$ an equation like:

$$\alpha_d T^d + \cdots + \alpha_0 + C(\beta_d T^d + \cdots + \beta_0) = 0$$

which expanded in μ gives rise to a linear system in α_i, β_j. We stress that [**Cou99**] gives an upper bound for the degree d.

Remark. The computation of the last two steps could involve computations with complex numbers, and so numerical approximations, because nothing tells us that the model we have chosen is defined over $\mathbb{Q}$. But, luckily, it is!

5. Numerical results.

The two covers we are looking for are given by:

$$\begin{array}{rccc} \varphi: & \mathcal{H} & \longrightarrow & \mathcal{M}_{0,4} \\ & T & \longmapsto & H_n(T) \end{array} \qquad \begin{array}{rccc} \Phi: & \mathcal{S} & \longrightarrow & \mathcal{M}_{0,5} \\ & X & \longmapsto & S_n(T,X). \end{array}$$

For $n = 6$, we obtain:

$$S_6(T,X) = \frac{\left(X^3 + \frac{75T+120}{16}X + \frac{625T^3+3600T^2+6720T+4096}{96T+256}\right)^2}{\frac{3(25T+56)^3}{2^8(3T+8)}\left(X^2 + TX + \frac{25T^3+120T^2+192T+128}{36T+96}\right)}$$

$$S_6(T,X) - 1 = \frac{(X-1)^3\left(X^3 + 3X^2 + \frac{75T+168}{8}X + \frac{625T^3+4950T^2+12960T+11136}{48T+128}\right)}{\frac{3(25T+56)^3}{2^8(3T+8)}\left(X^2 + TX + \frac{25T^3+120T^2+192T+128}{36T+96}\right)}$$

$$\begin{aligned} &S_6(T,X) - H_6(T) \\ &= \frac{\left(X^2 - \frac{5T+8}{2}X + \frac{125T^3+1050T^2+2720T+2176}{48T+128}\right)}{\frac{3(25T+56)^3}{2^8(3T+8)}\left(X^2 + TX + \frac{25T^3+120T^2+192T+128}{36T+96}\right)} \\ &\qquad \cdot \left(X^2 + \frac{5T+8}{4}X + \frac{25T^3+180T^2+424T+320}{24T+64}\right)^2 \end{aligned}$$

and:

$$H_6(T) = \frac{(T+8)(T+\frac{13}{5})^2(T+\frac{8}{5})^3}{-3\times 5(T+\frac{8}{3})(T+\frac{56}{25})^3}$$

$$H_6(T) - 1 = \frac{(T+2)(T+\frac{16}{5})^5}{-3\times 5(T+\frac{8}{3})(T+\frac{56}{25})^3}.$$

We manage similar computation in MAGMA for all n up to 20 in less than 20 minutes on an AMD 700Mhz.

In order to obtain totally real $\mathbb{Q}$-extension of degree n with Galois group $\mathcal{S}_n$ and $\mathcal{A}_n$, we have to specialize twice. The method points out a special value $t_h \in \mathbb{Q}$ of the parameter corresponding to the cover h we have deformed. Let us choose a close enough rational number $t_0 < t_h$ and set $t = t_0$. We get a regular $\mathcal{S}_n$-extension:

$$\begin{tikzcd} L = \mathbb{Q}(X)^{\mathrm{gal}} \arrow[dd, dash, "\mathcal{S}_n"'] \arrow[dr, dash] & \\ & \mathbb{Q}(X) \\ \mathbb{Q}(x) \arrow[ur, dash, "n"'] & \end{tikzcd} \qquad \text{where:} \quad x = S_n(t_0, X) \in \mathbb{Q}(X)$$

with four ramified points $0, H_n(t_0), 1$ and ∞ in this order. Moreover, due to the criterion of Section 3.1, all the points in $]0, H_n(t_0)[$ have totally real fibers. By Hilbert irreducibility theorem, most rationals in this interval provide totally real $\mathcal{S}_n$-extensions of $\mathbb{Q}$. The case of $\mathcal{A}_n$ follows classically.

References

[BF83] R. Biggers and M.D. Fried, *Moduli spaces of covers and the Hurwitz monodromy group*, Crelle's Journal, **335** (1983), 87-121, MR 84h:14034a, Zbl 0484.14002.

[Bir75] J.-S. Birman, *Braids, links and mapping class groups*, Ann. of Math. Studies, **82**, Princeton University Press, 1975, MR 51 #11477, Zbl 0305.57013.

[Cou99] J.-M. Couveignes, *Tools for the computation of families of coverings*, in 'Aspects of Galois theory' (H. Völkein, D. Harbater, P. Müller and J.G. Thompson, editors), London Math. Soc. Lect. Note Series, **256**, 38-65, Cambridge, 1999. (The following pages of this book were inadvertently numbered incorrectly: p. 21 should be p. 22, p. 22 should be p. 23 and p. 23 should be p. 21), MR 2000f:14037.

[Cou00] ______, *Boundary of Hurwitz spaces and explicit patching*, J. Symbolic Comp., **30**(**6**) (2000), 739-759, MR 2001m:12008, Zbl 0973.12004.

[CG94] J.-M. Couveignes and L. Granboulan, *Dessins from a geometric point of view*, in 'The Grothendieck Theory of Dessins d'Enfants' (L. Schneps, editor), London Math. Soc. Lect. Note Series, **200**, 79-113. Cambridge, 1994, MR 96b:14015, Zbl 0835.14010.

[DF94] P. Débes and M.D. Fried, *Nonrigid constructions in galois theory*, Pacific J. Math., **163**(**1**) (1994), 81-122, MR 95c:12008, Zbl 0788.12001.

[FD90] ______, *Rigidity and real residue class fields*, Acta Arithmetica, **LVI** (1990), 291-322, MR 92c:12005, Zbl 0667.14014.

[FV91] M.D. Fried and H. Völklein, *The inverse galois problem and rational points on moduli spaces*, Math. Ann., **290** (1991), 771-800, MR 93a:12004, Zbl 0763.12004.

[Ful69] W. Fulton, *Hurwitz schemes and irreducibility of moduli of algebraic curves*, Ann. of Math., **90** (1969), 542-575, MR 41 #5375, Zbl 0194.21901.

[GHP88] L. Gerritzen, F. Herrlich and M. van der Put, *Stable n-pointed trees of projective lines*, in 'Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen', **91**, 131-163, 1988, MR 89i:14005, Zbl 0698.14019.

[Han89] V.L. Hansen, *Braids and Coverings – Selected Topics*, London Math. Soc. Stud. Texts, **18**, Cambridge University Press, 1989, MR 94g:57004, Zbl 0692.57001.

[Har80] D. Harbater, *Deformation theory and the tame fundamental group*, Trans. Amer. Math. Soc., **262**(**2**) (1980), 399-415, MR 82f:14014, Zbl 0461.14001.

[Har87] ______, *Galois coverings of the arithmetic line*, Lect. Notes in Math., **1240**, Springer, 1987, 165-195, MR 88h:14020, Zbl 0627.12015.

[MM99] G. Malle and B.H. Matzat, *Inverse Galois Theory*, Springer, 1999, MR 2000k:12004, Zbl 0940.12001.

[Mes90] J.-F. Mestre, *Extensions régulières de $\mathbb{Q}(T)$ de groupe de Galois $\widetilde{A}_n$*, J. Algebra, **131** (1990), 483-495, MR 91j:12010, Zbl 0714.11074.

[Ser92] J.-P. Serre, *Topics in Galois Theory*, Jons and Barlett, 1992, MR 94d:12006, Zbl 0746.12001.

[Völ96] H. Völklein, *Groups as Galois Groups*, Cambridge Studies in Advanced Math., **53**, Cambridge, 1996, MR 98b:12003, Zbl 0868.12003.

[Wew99] S. Wewers, *Deformation of tame admissible covers of curves*, in 'Aspects of Galois theory' (H. Völkein, D. Harbater, P. Müller and J.G. Thompson, editors), London Math. Soc. Lect. Note Series, **256**, 239-282, Cambridge, 1999, MR 2001b:14048, Zbl 0995.14008.

Received February 7, 2002 and revised September 24, 2002.

Univerité de Toulouse II, le Mirail
Département Mathématiques - Informatique (Equipe G.R.I.M.M.)
5, Allées Antonio-Machado
31058 Toulouse Cedex 9
France
E-mail address: hallouin@univ-tlse2.fr

Univerité de Toulouse II, le Mirail
Département Mathématiques - Informatique (Equipe G.R.I.M.M.)
5, Allées Antonio-Machado
31058 Toulouse Cedex 9
France
E-mail address: riboulet@univ-tlse2.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

A PHRAGMÈN–LINDELÖF THEOREM AND THE BEHAVIOR AT INFINITY OF SOLUTIONS OF NON-HYPERBOLIC EQUATIONS

Zhiren Jin and Kirk Lancaster

We prove a Phragmèn–Lindelöf theorem which yields the behavior at infinity of bounded solutions of Dirichlet problems for non-hyperbolic (e.g., elliptic, parabolic) quasilinear second-order partial differential equations in terms of particular solutions of appropriate ordinary differential equations.

1. Introduction.

Many types of Phragmèn-Lindelöf Theorem have appeared in the literature since Edvard Phragmèn and Ernst Lindelöf's famous 1908 article ([**20**]; also see [**3**], Ch. 3). When Ω is an unbounded domain and $f \in C^2(\Omega) \cap C^0(\overline{\Omega})$ is a solution of a Dirichlet problem on Ω for a second-order elliptic or non-hyperbolic equation, a fundamental question is that of the behavior of $f(X)$ as $|X|$ goes to infinity. A Phragmèn-Lindelöf theorem "at infinity" establishes the existence of asymptotic limits of f at infinity and offers insight into the nature of these limits when f lies in an appropriate class of solutions. The goal of this note is to obtain a comprehensive Phragmèn-Lindelöf theory at infinity for bounded solutions of Dirichlet problems in certain types of domains using the "local barrier functions" constructed in [**14**] and solutions of boundary value problems for ordinary differential equations.

Let Ω be an open set in $\mathbf{R}^n$. Suppose $(a_{ij}(X, z, P))$ is any $n \times n$ (symmetric) matrix with trace one which is positive semidefinite for $X \in \Omega$, $z \in \mathbf{R}$, and $P \in \mathbf{R}^n$ and whose entries satisfy $a_{ij} \in C^0(\Omega \times \mathbf{R} \times \mathbf{R}^n)$. Assume further that $a_{nn}(X, z, P) \geq \sigma_1(|P|)$ for some positive continuous function σ_1 defined on $[1, \infty)$. Let b be a function in $C^0(\Omega \times \mathbf{R} \times \mathbf{R}^n)$. Let Q be the non-hyperbolic operator defined by

$$(1) \quad Qu(X) = \sum_{i,j=1}^{n} a_{ij}(X, u(X), Du(X))D_{ij}u(X) + b(X, u(X), Du(X)).$$

For convenience, let us write elements $X = (x_1, \ldots, x_n)$ of $\mathbf{R}^n$ as $(\mathbf{x}, y)$, where

$$\mathbf{x} = (x_1, \ldots, x_{n-1}) \quad \text{and} \quad y = x_n$$

and, for each $M > 0$, let S_M denote the set

$$\{X = (x_1, \ldots, x_n) \in \mathbf{R}^n \mid |x_n| < M\}.$$

Let us assume here that $\Omega \subset S_M$ for some $M > 0$. If b/a_{nn} have appropriate limits at infinity (i.e., (7)), $\phi \in C^0(\mathbf{R}^n)$, $\omega \in S^{n-2}$ is a direction of Ω at infinity (i.e., (6)) and Assumptions 1 and 2 in §2 are satisfied, we will prove that every bounded solution $f \in C^2(\Omega) \cap C^0(\overline{\Omega})$ of the Dirichlet problem

$$Qf = 0 \quad \text{in } \Omega \tag{2}$$

and

$$f = \phi \quad \text{on } \partial\Omega \tag{3}$$

satisfies

$$f(\mathbf{x}, y) \to k_\omega(y) \qquad \text{for} \quad X = (\mathbf{x}, y) \in \overline{\Omega} \tag{4}$$

as $|X| \to \infty$ with $\frac{\mathbf{x}}{|\mathbf{x}|} \to \omega$, where k_ω is a solution of a related boundary value problem (e.g., (13)).

We will use local barrier functions and solutions of ordinary differential equations to obtain (4). Rescaling (and truncating) the graph of a barrier function w while leaving unaltered a solution k of an appropriate ordinary differential equation and comparing a bounded solution f of the Dirichlet problem with $w + k$ is a principal technique we will use. As a consequence of the facts that our fundamental comparisons are made in barrier domains of the form $U = \{X \in \mathbf{R}^n \mid C_1 < X \cdot \nu < C_2, |X - (X \cdot \nu)\nu| < h(X \cdot \nu)\}$ for $\nu \in S^{n-1}$ and $C_1 < C_2$ and the ability to rescale w (and so h) improves our estimates, domains in slabs are of particular interest. Our results are significant for:

(i) The generality of allowable domains Ω,
(ii) the generality of allowable operators Q, and
(iii) the simplification achieved by approximating $f(\mathbf{x}, y)$ by $k(y)$ for $|\mathbf{x}|$ large when f is an "unknown" solution of (2) & (3) and k is a "known" solution of a boundary value problem for an ordinary differential equation (e.g., (13)).

Theorem 2.2 complements other Phragmèn-Lindelöf principles at infinity in which the domain has different geometric constraints, for example being required to lie in a cone (e.g., [**1**], [**17**]). The results here hold for a large class of operators, including uniformly elliptic operators, degenerate elliptic operators and parabolic operators. These results can be used to investigate other questions, such as the effect on the behavior at infinity of a solution f when the coefficients of Q are perturbed. Finally, the approximation (near infinity) of the solution of a partial differential equation by the solution of an ordinary differential equation (i.e., (iii)) is a very useful technique which is often used, sometimes without justification, in continuum mechanics (e.g., [**9**]).

Previous results on Phragmèn-Lindelöf theorems at infinity generally concern limited classes of operators and/or limited types of domains. The cases in which Ω is (or is contained in) a strip in $\mathbf{R}^2$ or a cylinder in $\mathbf{R}^3$ have generated particular interest, in part because of applications of Phragmèn-Lindelöf principles and their companion "spatial decay estimates" to problems in continuum mechanics (e.g., [**7**]; also see references in [**14**], [**15**]). Classes of operators in previous articles include (linear and nonlinear) uniformly elliptic operators or divergence structure operators (e.g., [**1**], [**2**], [**11**], [**12**], [**21**]). Theorems containing decay estimates usually concern limited classes of operators in special geometries, including strips (e.g., [**10**], [**11**]) and cylinders (e.g., [**2**], [**8**]).

In [**14**], Dirichlet problems in domains $\Omega \subset S_M$ for quasilinear elliptic second-order partial differential equations which do not have lower-order terms are studied. It is shown there that $f(\mathbf{x}, y) \to \Phi(\omega)$ as $|\mathbf{x}| \to \infty$ with $\frac{\mathbf{x}}{|\mathbf{x}|} \to \omega$ when ω is a "direction of Ω at infinity," f is a solution of the Dirichlet problem with Dirichlet data ϕ, $\phi(\mathbf{x}, y) \to \Phi(\omega)$ as $|\mathbf{x}| \to \infty$ with $\frac{\mathbf{x}}{|\mathbf{x}|} \to \omega$, and $b \equiv 0$. In addition, if $\Omega = S_M$, $\phi(\mathbf{x}, M) \to \Phi_1(\omega)$ and $\phi(\mathbf{x}, -M) \to \Phi_2(\omega)$ as $|\mathbf{x}| \to \infty$ with $\frac{\mathbf{x}}{|\mathbf{x}|} \to \omega$, and the other conditions remain unchanged, it is proven that

$$f(\mathbf{x}, y) \to \frac{1}{2M}\big(\Phi_1(\omega) - \Phi_2(\omega)\big)(y + M) + \Phi_2(\omega)$$

as $|\mathbf{x}| \to \infty$ with $\frac{\mathbf{x}}{|\mathbf{x}|} \to \omega$. The results in [**14**] are significant for the generality of operators Q and domains Ω allowed and especially for the construction of new barrier functions. The inclusion of lower-order terms here complicates the arguments used in [**14**] in a subtle but significant way; we compensate for this in the Proof of Theorem 2.2 by assuming our solutions are bounded. All arguments occurring here are "local" with respect to the direction ω.

2. Main result.

We will assume from now on that the coefficients of Q have been normalized so that

$$\sum_{i=1}^{n} a_{ii}(X, z, P) = 1 \quad \text{for} \quad (X, z, P) \in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R}^n \tag{5}$$

and satisfy the conditions mentioned previously (i.e., before (1)). We will set $I_M = (-M, M)$ and

$$\pi(\Omega) = \{(x_1, \ldots, x_{n-1}) : \exists_{y \in [-M, M]}\ (x_1, \ldots, x_{n-1}, y) \in \Omega\}.$$

Let $T(\Omega)$ represent the set of directions $\omega \in S^{n-2}$ at infinity of Ω (actually $\pi(\Omega)$); that is

$$T(\Omega) = \cap_{N=1}^{\infty} \overline{\cup_{r \geq N} \{\omega \in S^{n-2} : r\omega \in \pi(\Omega)\}}. \tag{6}$$

Notice that $\omega \in T(\Omega)$ if and only if there exists a sequence $\{(\mathbf{x_j}, y_j)\}$ in Ω with $|\mathbf{x_j}| \to \infty$ and $\frac{\mathbf{x_j}}{|\mathbf{x_j}|} \to \omega$ as $j \to \infty$.

For $\omega \in T(\Omega)$, consider the following assumptions:

Assumption 1. For some open subset O of S^{n-2} with $\omega \in O$, there exists a function $E \in C^0(O \times I_M \times \mathbf{R}^2)$ such that $E\left(\frac{\mathbf{x}}{|\mathbf{x}|}, y, z, q\right)$ is nonincreasing in z and

$$\frac{b(\mathbf{x}, y, z, \mathbf{p}, q)}{a_{nn}(\mathbf{x}, y, z, \mathbf{p}, q)} \to E(\sigma, y, z, q) \tag{7}$$

as $|\mathbf{x}| \to \infty$ with $\frac{\mathbf{x}}{|\mathbf{x}|} \to \sigma$ and $|\mathbf{p}| \to 0$ uniformly for $|y| < M$, $\sigma \in O$, and $z, q \in \mathbf{R}$.

Assumption 2. There exists a function k mapping $\overline{I_M} \times T$ into $\mathbf{R}$ such that

$$\phi(\mathbf{x}, y) \to k(y, \omega) \tag{8}$$

uniformly as $|\mathbf{x}| \to \infty$ and $\frac{\mathbf{x}}{|\mathbf{x}|} \to \omega$ for $(\mathbf{x}, y) \in \partial\Omega$ and, for each $\alpha > 0$, there exist $\delta = \delta_{\alpha,\omega} > 0$ and functions k_1 and k_2 in $C^1(\overline{I_M}) \cap C^2(I_M)$ such that for each $y \in I_M$

$$|k_1(y) - k(y, \omega)| \leq \alpha, \tag{9}$$

$$|k_2(y) - k(y, \omega)| \leq \alpha, \tag{10}$$

$$k_1''(y) + E(\omega, y, k_1(y), k_1'(y)) \geq \delta, \tag{11}$$

and

$$k_2''(y) + E(\omega, y, k_2(y), k_2'(y)) \leq -\delta. \tag{12}$$

Remark 2.1. In what might be the most common situation in which Assumptions 1 and 2 are satisfied for all $\omega \in T(\Omega)$, we would have $\Omega = U \times I_M$ for some open subset U of $\mathbf{R}^{n-1}$, $E \in C^0(S^{n-2} \times I_M \times \mathbf{R}^2)$, $k \in C^0(\overline{I_M} \times T(\Omega))$, $k_\omega \in C^2(I_M)$,

$$k_\omega''(y) + E(\omega, y, k_\omega(y), k_\omega'(y)) = 0 \quad \text{for} \quad |y| < M, \omega \in T(\Omega), \tag{13}$$

where k_ω is defined by $k_\omega(y) = k(y, \omega)$ for $\omega \in T(\Omega)$, and, for each $\omega \in T(\Omega)$, functions k_1 and k_2 respectively satisfying (9)-(12).

Theorem 2.2. *Let $M > 0$, $\Omega \subset S_M$, and $\omega \in T(\Omega)$. Suppose:*

1) *$f \in C^2(\Omega) \cap C^0(\overline{\Omega}) \cap L^\infty(\Omega)$ satisfies* (2) & (3);
2) *Assumptions* 1 *and* 2 *are satisfied for ω*;

3) *there exist* $L \geq 0$ *and a positive continuous function* σ_1 *on* $[1, \infty)$ *such that*

$$a_{nn}(\mathbf{x}, y, z, \mathbf{p}, q) \geq \sigma_1(|\mathbf{p}|^2 + q^2) \tag{14}$$

whenever $\mathbf{x}, \mathbf{p} \in \mathbf{R}^{n-1}$, $y, z, q \in \mathbf{R}$ *with* $|\mathbf{x}| \geq L$ *and* $|y| < M$;

4) Q *satisfies* (5).

Then

$$\lim_{j \to \infty} |f(\mathbf{x}_j, y_j) - k(y_j, \omega)| = 0 \tag{15}$$

uniformly for sequences $\{(\mathbf{x}_j, y_j)\}$ *in* $\overline{\Omega}$ *with* $|\mathbf{x}_j| \to \infty$ *and* $\frac{\mathbf{x}_j}{|\mathbf{x}_j|} \to \omega$ *as* $j \to \infty$.

When Q is of a particular type (e.g., uniformly elliptic), arguments exist which show that a solution f of (2) & (3) is bounded whenever it satisfies an appropriate (for Q) growth condition. For such operators, we may assume that the hypothesis $f \in L^\infty(\Omega)$ in Theorem 2.2 is replaced by this growth condition without changing the conclusion of the theorem. From the Proof of Theorem 2.2, it follows that $f \in L^\infty(\Omega)$ can be replaced by f is "bounded in the direction ω" in the sense that there exist $\delta > 0$, $R > 0$, and $J \geq 0$ such that $|f(\mathbf{x}, y)| \leq J$ if $(\mathbf{x}, y) \in \overline{\Omega}$, $|\mathbf{x}| \geq R$, and $\left| \frac{\mathbf{x}}{|\mathbf{x}|} - \omega \right| < \delta$. Finally, the necessity of the nondegerancy condition on a_{nn} (i.e., (14)) is illustrated by Example 4 of [**15**].

We shall also prove the following consequence of Theorem 2.2:

Theorem 2.3. *Let* $\omega \in T$. *Suppose:*

1) $f \in C^2(\Omega) \cap C^0(\overline{\Omega}) \cap L^\infty(\Omega)$ *satisfies* (2) & (3);
2) *Assumption* 1 *is satisfied for* ω;
3) $E = E(y, z, q)$ *is a nonincreasing function of* z *for each* $(y, q) \in \overline{I} \times \mathbf{R}$;
4) E, $\frac{\partial E}{\partial z}$, $\frac{\partial E}{\partial q} \in C^0(\overline{I} \times \mathbf{R}^2)$;
5) *there exists* $k \in C^2(\overline{I_M})$ *such that*

$$k''(y) + E(\omega, y, k(y), k'(y)) = 0 \qquad \textit{for} \quad |y| < M$$

and $\phi(\mathbf{x}, y) \to k(y)$ *uniformly as* $|\mathbf{x}| \to \infty$ *and* $\frac{\mathbf{x}}{|\mathbf{x}|} \to \omega$ *for* $(\mathbf{x}, y) \in \partial\Omega$;

6) *there exist* $L \geq 0$ *and a positive continuous function* σ_1 *on* $[1, \infty)$ *such that*

$$a_{nn}(\mathbf{x}, t, z, \mathbf{p}, q) \geq \sigma_1(|\mathbf{p}|^2 + q^2)$$

whenever $\mathbf{x}, \mathbf{p} \in \mathbf{R}^{n-1}$, $z, t, q \in \mathbf{R}$ *with* $|\mathbf{x}| \geq L$ *and* $|t| \leq M$;

7) Q *satisfies* (5).

Then

$$\lim_{j \to \infty} |f(\mathbf{x}_j, y_j) - k(y_j, \omega)| = 0 \tag{16}$$

uniformly for sequences $\{(\mathbf{x}_j, y_j)\}$ *in* $\overline{\Omega}$ *with* $|\mathbf{x}_j| \to \infty$ *and* $\frac{\mathbf{x}_j}{|\mathbf{x}_j|} \to \omega$ *as* $j \to \infty$.

3. Barrier functions.

Let us review the construction of barrier functions in [**14**]. This idea originated from the following fact: If $w = w(\mathbf{x}, y)$ in $C^2(\mathbf{R}^n)$ satisfies $Qw = 0$, Q is elliptic (non-hyperbolic), and $g = g(\mathbf{x}, z)$ is a function in $C^2(\mathbf{R}^n)$ for which $g_z \neq 0$ and

$$g(\mathbf{x}, w(\mathbf{x}, y)) = y,$$

then g will satisfy an equation of the form $Q^{\#}g = 0$ for an elliptic (respectively non-hyperbolic) operator $Q^{\#}$, where Q and $Q^{\#}$ are related by the equation

$$Qw(\mathbf{x}, y) = \frac{-1}{g_z^3(\mathbf{x}, w(\mathbf{x}, y))} Q^{\#} g(\mathbf{x}, w(\mathbf{x}, y)). \tag{17}$$

In particular, if $g_z > 0$ and $Q^{\#}g > 0$, then $Qw < 0$. A computation shows that $Q^{\#}$ is defined by

$$Q^{\#}v(\mathbf{x}, z) = \sum_{i,j=1}^{n} A_{ij}(\mathbf{x}, z, v, Dv) D_{ij} v + B(\mathbf{x}, z, v, Dv) \tag{18}$$

for $v = v(\mathbf{x}, z)$ in $C^2(R^n)$ with $\frac{\partial v}{\partial z} \neq 0$, where

$$A_{ij}(\mathbf{x}, z, t, \mathbf{p}, q) = q^2 a_{ij}, \qquad 1 \leq i, j \leq n-1, \tag{19}$$

$$A_{in}(\mathbf{x}, z, t, \mathbf{p}, q) = q a_{in} - \sum_{j=1}^{n-1} p_j q a_{ij}, \qquad 1 \leq i \leq n-1, \tag{20}$$

$$A_{nn}(\mathbf{x}, z, t, \mathbf{p}, q) = a_{nn} - 2\sum_{j=1}^{n-1} p_j a_{jn} + \sum_{i,j=1}^{n-1} p_i p_j a_{ij}, \tag{21}$$

and

$$B(\mathbf{x}, z, t, \mathbf{p}, q) = -q^3 b\left(\mathbf{x}, t, z, -\frac{\mathbf{p}}{q}, \frac{1}{q}\right). \tag{22}$$

Here a_{ij} means $a_{ij}(\mathbf{x}, t, z, -\frac{\mathbf{p}}{q}, \frac{1}{q})$ for $1 \leq i, j \leq n$, $\mathbf{p} = (p_1, \ldots, p_{n-1}) \in \mathbf{R}^{n-1}$, $t \in \mathbf{R}$, $q \neq 0$, $D_i = \frac{\partial}{\partial x_i}$ for $1 \leq i \leq n-1$, $D_n = \frac{\partial}{\partial z}$, $Dv = (D_1 v, \ldots, D_n v)$, and $D_{ij} = D_i D_j$ for $1 \leq i, j \leq n$. The construction of barriers for Q is somewhat similar to the constructions of barriers for the operator $Q^{\#}$ given in [**13**] and [**22**].

In this note, we will be unable to use barriers specifically tailored to our operator as was done in [**14**]. Instead, in the construction in §7 of [**14**], we will set $\sigma \equiv 1$ and obtain the functions

$$\chi(\alpha) = \begin{cases} \frac{1}{2} - \ln(\alpha) & \text{if } 0 < \alpha < 1 \\ \frac{1}{2\alpha^2} & \text{if } 1 \le \alpha < \infty \end{cases}$$

and

$$\eta(\beta) = \begin{cases} \frac{1}{\sqrt{2\beta}} & \text{if } 0 < \beta < \frac{1}{2} \\ e^{\frac{1}{2}-\beta} & \text{if } \frac{1}{2} \le \beta < \infty. \end{cases}$$

Also let us define for $H \ge 1$ the number

$$A(H) = 2M \left(\int_1^{e^{\chi(H)}} \eta(\ln(t))\, dt \right)^{-1}.$$

Then for $a > 0$, $H \ge 1$, $\mathbf{x}_0 \in \mathbf{R}^{n-1}$, and $\Gamma \in \mathbf{R}$, the construction in §7 of [**14**] yields the functions $h_a = h_{a,H}$, $g_a = g_{a,\mathbf{x}_0,\Gamma,M,H}$, and $w_a = w_{a,\mathbf{x}_0,\Gamma,H}$ defined by

$$h_a(r) = \begin{cases} a\sqrt{e}\left(\frac{1}{2H^2} - \frac{1}{2} \right) + \frac{a}{\sqrt{2}}\left(\lambda(\sqrt{e}) - \lambda\left(\frac{r}{a}\right)\right) & \text{if } a < r < a\sqrt{e} \\ a\sqrt{e}\left(\frac{1}{2H^2} - \ln\left(\frac{r}{a}\right) \right) & \text{if } a\sqrt{e} \le r < ae^{\chi(H)}, \end{cases}$$

$$g_a(\mathbf{x}, z) = h_a\left(\sqrt{|\mathbf{x} - \mathbf{x}_0|^2 + (z - \Gamma)^2}\right) - M,$$

and

$$w_a(\mathbf{x}, y) = \Gamma - \sqrt{(h_a^{-1}(y + M))^2 - |\mathbf{x} - \mathbf{x}_0|^2},$$

where λ satisfies $\lambda'(t) = \frac{1}{\sqrt{\ln(t)}}$.

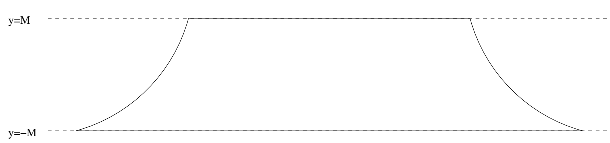


Figure 1. $\Omega_{a,\mathbf{x}_0,H}$.

The domain of $w_{a,\mathbf{x}_0,\Gamma,H}$ is a set $\Omega_{a,\mathbf{x}_0,H} \subset S_M$, illustrated in Figure 1 when $n = 2$, which is relatively compact in $\mathbf{R}^n$ and whose central axis (of symmetry) is $\{(\mathbf{x}_0, y) : |y| < M\}$. As the parameter a becomes larger, the domain $\Omega_{a,\mathbf{x}_0,H}$ becomes larger but the variation of w_a along the axis of symmetry decreases and goes to zero as a goes to infinity. This "rescaling" of the barrier w_a by increasing a allows increasingly better estimates of a solution along the central axis; this fact plays a key role in the use of these barriers. As a goes to infinity, w_a also goes to infinity on $\partial\Omega_{a,\mathbf{x}_0,H} \cap S_M$. We assume a solution f is bounded in order to use this fact to help show that $f \le w_a + k_2$ in the Proof of Theorem 2.2; a careful examination of the growth rate of w_a on $\partial\Omega_{a,\mathbf{x}_0,H} \cap S_M$ might allow the growth hypothesis on f (i.e., f is bounded) to be relaxed (e.g., Theorem 2.5 of [**14**]).

4. Proofs of Theorems 2.2 & 2.3.

Proof of Theorem 2.2. We may assume that the set O mentioned in Assumption 1 is all of S^{n-2}. As described in the previous section, in the construction in §7, [**14**] we let $\sigma \equiv 1$ (we ignore (7.1), [**14**]), so that $\Psi \equiv 1$. Let $\omega \in T$, $\epsilon > 0$, and $\alpha = \epsilon$. Let $\delta = \delta_{\alpha,\omega}$, k_1 and k_2 be as given in Assumption 2. From Assumption 2 and the continuity of $k(y,\omega)$, we see that there exist $\delta_1 > 0$ and R_1 such that if $(\mathbf{x}, y) \in \partial\Omega$, $|\mathbf{x}| \ge R_1$, $|y| \le M$, and $\left|\frac{\mathbf{x}}{|\mathbf{x}|} - \omega\right| < \delta_1$, we have

$$|\phi(\mathbf{x}, y) - k(y,\omega)| < \epsilon. \tag{23}$$

Assumption 1 implies there exist $\delta_2 > 0$ and R_2 such that

$$\left|\frac{b(\mathbf{x}, y, z, \mathbf{p}, q)}{a_{nn}(\mathbf{x}, y, z, \mathbf{p}, q)} - E(\omega, y, z, q)\right| \le \frac{\delta}{4} \tag{24}$$

if $|\mathbf{x}| \ge R_2$, $|\mathbf{p}| \le \delta_2$, and $\left|\frac{\mathbf{x}}{|\mathbf{x}|} - \omega\right| \le 2\delta_2$. Consider the compact set

$$K = \{(\mathbf{p}, q) \in \mathbf{R}^n : |\mathbf{p}|^2 + q^2 \le 2(1 + \|k_2'\|_\infty^2)\}.$$

From (14), we see that there exists $\mu(K) > 0$ such that

$$a_{nn}(\mathbf{x}, y, z, \mathbf{p}, q) \ge \mu(K) \qquad \text{if} \quad (\mathbf{p}, q) \in K, \mathbf{x} \in \mathbf{R}^{n-1} \quad \text{and} \quad y, z \in \mathbf{R}.$$

Set $N = \|f - k_2\|_\infty$. Since for fixed ω, $E(\omega, y, z, q)$ is uniformly continuous for (y, z, q) in a fixed compact set, there exists $\delta_3 > 0$ such that

$$|E(\omega, y, z, t + q) - E(\omega, y, z, q)| \le \frac{\delta}{4} \tag{25}$$

for $|t| \le \delta_3$, $|y| < M$, $|z| < N$, and $|q|^2 \le 2(1 + \|k_2'\|_\infty^2)$. Let us set $\delta_0 = \min\{1, \delta_1, \delta_2, \delta_3\}$ and choose $H \ge 2$ such that $\frac{2M}{H} < \epsilon$, $\chi(H) \le \ln(2)$,

$$A(H) \ge 16N, \qquad A(H) \ge \frac{5}{\mu(K)\delta}, \tag{26}$$

and

$$\frac{2\sqrt{2NA(H)e^{\chi(H)}}}{A(H)} + \frac{2}{H} < \delta_0, \tag{27}$$

where $A(H)$ is given in (7.8), [**14**]. There exists $R_3 > 0$ such that if $|\mathbf{x}_0| \ge R_3$, $|\mathbf{x} - \mathbf{x}_0| \le A(H)e^{\chi(H)}$, and $\left|\frac{\mathbf{x}_0}{|\mathbf{x}_0|} - \omega\right| < \delta_0$, then $\left|\frac{\mathbf{x}}{|\mathbf{x}|} - \omega\right| < 2\delta_0$. Set $R_0 = \max\{R_1, R_2, R_3\} + A(H)e^{\chi(H)}$.

Now define

$$W = \left\{ \mathbf{x} \mid |\mathbf{x}| > R_0, \quad \left| \frac{\mathbf{x}}{|\mathbf{x}|} - \omega \right| < \delta_0 \right\}.$$

We claim that if $(\mathbf{x}_0, y) \in \overline{\Omega}$ and $\mathbf{x}_0 \in W$, then

$$f(\mathbf{x}_0, y) < k(y, \omega) + 2\epsilon. \tag{28}$$

Throughout the remainder of this proof, let $\mathbf{x}_0$ represent a point in W such that $(\mathbf{x}_0, y) \in \overline{\Omega}$ for some $y \in I_M$.

Let $w(\mathbf{x}, y) = w_{a,\mathbf{x}_0,\gamma,H}(\mathbf{x}, y)$ be the upper barrier given by (7.14), [**14**] with $\gamma = 2\epsilon$ and $a = A(H)$; a formula for w_a is given in the previous section. Notice then that $w \geq \gamma = 2\epsilon$ on $\Omega_{a,\mathbf{x}_0,H}$. Now set

$$\Omega_1 = \left\{ (\mathbf{x}, y) \in \Omega_{a,\mathbf{x}_0,H} \cap \Omega \ : \ |\mathbf{x} - \mathbf{x}_0| < \sqrt{2NA(H)e^{\chi(H)} - N^2} \right\}, \tag{29}$$

which is illustrated by the shaded region in Figure 2 when $n = 2$,

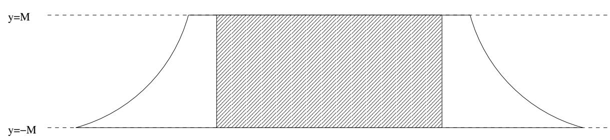


Figure 2. Ω_1.

and define $u_2 \in C^1(\overline{\Omega_1}) \cap C^2(\Omega_1)$ by

$$u_2(\mathbf{x}, y) = w(\mathbf{x}, y) + k_2(y).$$

Notice that if $(\mathbf{x}, y) \in \Omega_1$, then $|\mathbf{x}| \geq \max\{R_1, R_2, R_3\}$, $h'_a(h_a^{-1}(y + M)) \geq H \geq 2$, $A(H) < h_a^{-1}(y + M) < A(H)e^{\chi(H)}$, and $\left| \frac{\mathbf{x}}{|\mathbf{x}|} - \omega \right| < 2\delta_0$.

Let $\zeta \geq 0$. We claim that

$$Q(u_2 + \zeta) < 0 \ \text{ in } \ \Omega_1. \tag{30}$$

From §7, [**14**], we find that

$$\frac{\partial^2 w}{\partial x_i \partial x_j}(\mathbf{x}, y) = \frac{\delta_{ij} S^2 + (x_i - x_i^{(0)})(x_j - x_j^{(0)})}{S^3} \quad \text{for} \quad 1 \leq i, j \leq n - 1,$$

$$\frac{\partial^2 w}{\partial x_i \partial y}(\mathbf{x}, y) = \frac{-(x_i - x_i^{(0)}) Z}{S^3 h'_a(Z)} \quad \text{for} \quad 1 \leq i \leq n - 1,$$

$$\frac{\partial^2 w}{\partial y^2}(\mathbf{x}, y) = \frac{S^2 \big(Z\ h''_a(Z) - h'_a(Z) \big) + Z^2 h'_a(Z)}{S^3 (h'_a(Z))^3}$$

where $\mathbf{x}_0 = \left(x_1^{(0)}, \ldots, x_{n-1}^{(0)}\right)$,

$$Z = h_a^{-1}(y+M), \qquad \text{and}$$
$$S = \sqrt{(h_a^{-1}(y+M))^2 - |\mathbf{x}-\mathbf{x}_0|^2} = \sqrt{Z^2 - |\mathbf{x}-\mathbf{x}_0|^2}.$$

Since $A(H) < Z < 2A(H)$, $|\mathbf{x}-\mathbf{x}_0|^2 < 2NA(H)e^{\chi(H)} - N^2 \le 4NA(H) - N^2$, and $A(H) \ge 16N$, it is easy to see that $2S^2 \ge (A(H))^2$. Notice then that

$$|Dw(\mathbf{x},y)| \le \frac{|\mathbf{x}-\mathbf{x}_0|}{S} + \frac{Z}{S|h_a'(Z)|} \le \frac{2\sqrt{2NA(H)e^{\chi(H)} - N^2}}{A(H)} + \frac{2}{H}$$

and so (27) implies $|Dw(\mathbf{x},y)| < \delta_0$.

If we set $\xi_i = \frac{x_i - x_i^{(0)}}{S}$ for $1 \le i \le n-1$, $\xi_n = \frac{-Z}{Sh_a'(Z)}$, and $\xi = (\xi_1, \ldots, \xi_n)$, then $|\xi| \le 1$ and

$$\frac{1}{S}\sum_{i,j=1}^{n} a_{ij}(\mathbf{x},y,u_2+\zeta,Du_2)\xi_i\xi_j \le \frac{1}{S}.$$

Since

$$\sum_{i,j=1}^{n-1} a_{ij}(\mathbf{x},y,u_2+\zeta,Du_2)\frac{\delta_{ij}}{S} = \frac{1}{S}(1 - a_{nn}(\mathbf{x},y,u_2+\zeta,Du_2))$$

and

$$\frac{Zh_a''(Z)}{S(h_a'(Z))^3} = -\frac{1}{S},$$

we have

$$\sum_{i,j=1}^{n} a_{ij}(\mathbf{x},y,u_2+\zeta,Du_2)D_{ij}w(\mathbf{x},y)$$
$$\le \frac{2}{S} - \frac{1}{S}\left(2 + \frac{1}{(h_a'(Z))^2}\right)a_{nn}(\mathbf{x},y,u_2+\zeta,Du_2) < \frac{2}{S}.$$

Since $|Dw(\mathbf{x},y)| < \delta_0$ when $(\mathbf{x},y) \in \Omega_1$, we have $Du_2(\mathbf{x},y) \in K$ and so $a_{nn}(\mathbf{x},y,u_2+\zeta,Du_2) \ge \mu(K)$ if $(\mathbf{x},y) \in \Omega_1$. Set $\mu = \mu(K)$. From (26), we obtain $\frac{2}{S} \le \frac{\mu\delta}{2}$. Notice that

$$E(\omega,y,u_2+\zeta,q) \le E(\omega,y,u_2,q) \le E(\omega,y,k_2,q) \tag{31}$$

for all $y \in I_M$ and $q \in \mathbf{R}$ since $\zeta \ge 0$ and $u_2 = w + k_2 \ge 2\epsilon + k_2 > k_2$. Using (24), (25) and (31), we have

$$Q(w+k_2+\zeta)(\mathbf{x},y)$$
$$= \sum_{i,j=1}^{n-1} a_{ij}D_{ij}w + 2\sum_{i=1}^{n-1} a_{in}D_{in}w + (D_{nn}w + k_2''(y))a_{nn} + b$$

$$
\begin{aligned}
&< \frac{\mu\delta}{2} + \left(k_2''(y) + \frac{b(u_2+\zeta, Du_2)}{a_{nn}(u_2+\zeta, Du_2)} \right) a_{nn}(u_2+\zeta, Du_2) \\
&= \frac{\mu\delta}{2} + \left[\frac{b(u_2+\zeta, Du_2)}{a_{nn}(u_2+\zeta, Du_2)} - E(u_2+\zeta, D_n u_2) \right. \\
&\quad + E(u_2+\zeta, D_n u_2) - E(u_2, D_n u_2) + E(u_2, D_n u_2) - E(u_2, k_2') \\
&\quad \left. + E(u_2, k_2') - E(k_2, k_2') + E(k_2, k_2') + k_2''(y) \right] a_{nn} \\
&\le \frac{\mu\delta}{2} + \left[\frac{\delta}{4} + \frac{\delta}{4} - \delta \right] a_{nn}(Du_2) \\
&\le \frac{\mu\delta}{2} - \frac{\mu\delta}{2} = 0,
\end{aligned}
$$

where we abbreviate $a_{nn} = a_{nn}(u_2+\zeta, Du_2) = a_{nn}(\mathbf{x}, y, u_2+\zeta, Du_2)$, $b = b(u_2+\zeta, Du_2) = b(\mathbf{x}, y, u_2+\zeta, Du_2)$, $E(u_2, D_n u_2) = E(\omega, y, u_2, D_n u_2) = E(\omega, y, u_2, D_n w + k_2'(y))$, $E(u_2, k_2'(y)) = E(\omega, y, u_2, k_2'(y))$ and $E(k_2(y), k_2'(y)) = E(\omega, y, k_2(y), k_2'(y))$.

If $(\mathbf{x}, y) \in \partial\Omega \cap \partial\Omega_1$, from (7.3), [**14**] and (23) we have

$$f(\mathbf{x}, y) = \phi(\mathbf{x}, y) < k_\omega(y) + \epsilon \le k_2(y) + 2\epsilon = k_2(y) + \gamma \le w(\mathbf{x}, y) + k_2(y).$$

Thus

$$f(\mathbf{x}, y) - u_2(\mathbf{x}, y) < 0 \text{ on } \partial\Omega \cap \partial\Omega_1.$$

If $(\mathbf{x}, y) \in \Omega \cap \partial\Omega_1$, then $|\mathbf{x} - \mathbf{x}_0| = \sqrt{2NA(H)e^{\chi(H)} - N^2}$ and so

$$
\begin{aligned}
w(\mathbf{x}, y) &= 2\epsilon + A(H)e^{\chi(H)} - \sqrt{\left(h_a^{-1}(y+M)\right)^2 - |\mathbf{x} - \mathbf{x}_0|^2} \\
&\ge 2\epsilon + A(H)e^{\chi(H)} \\
&\quad - \sqrt{\left(A(H)e^{\chi(H)}\right)^2 - 2NA(H)e^{\chi(H)} + N^2} \\
&= 2\epsilon + N.
\end{aligned}
$$

Hence

$$f(\mathbf{x}, y) - k_2(y) \le \|f - k_2\|_\infty = N \le w(\mathbf{x}, y) - 2\epsilon < w(\mathbf{x}, y)$$

and so $f(\mathbf{x}, y) < u_2(\mathbf{x}, y)$ for $(\mathbf{x}, y) \in \Omega \cap \partial\Omega_1$.

Let $U_0 = \{(\mathbf{x}, y) \in \Omega_1 \ : \ f(\mathbf{x}, y) > u_2(\mathbf{x}, y)\}$. Since $f < u_2$ on $\partial\Omega_1$, U_0 is a relatively compact subset of Ω_1 and $f = u_2$ on $\Omega_1 \cap \partial U_0$. Now define

$$Ru(\mathbf{x}, y) = \sum_{ij=1}^{n} \bar{a}_{ij}(\mathbf{x}, y, Du) D_{ij} u(\mathbf{x}, y) + \bar{b}(\mathbf{x}, y, Du)$$

by setting $\bar{a}_{ij}(\mathbf{x}, y, q) = a_{ij}(\mathbf{x}, y, f(\mathbf{x}, y), q)$ and $\bar{b}(\mathbf{x}, y, q) = b(\mathbf{x}, y, f(\mathbf{x}, y), q)$. Let $(\mathbf{x}_1, y_1)$ be an arbitrary point in U_0 and set $\zeta = f(\mathbf{x}_1, y_1) - u_2(\mathbf{x}_1, y_1) > 0$.

Since $Q(u_2+\zeta)<0$ on Ω_1, we have

$$Ru_2(\mathbf{x}_1,y_1)=Q(u_2+\zeta)(\mathbf{x}_1,y_1)<0.$$

Since $(\mathbf{x}_1,y_1)$ is an arbitrary point in U_0, we have $Ru_2<0$ in U_0. Recalling that the ellipticity of R is not needed in Theorem 10.1 of [**4**] (as noted in the proof of Theorem 3.1 of [**4**]), we see that $f\le u_2$ on U_0. Hence $U_0=\emptyset$ and so

$$f(\mathbf{x},y)\le u_2(\mathbf{x},y) \quad\text{on}\quad \Omega_1.$$

Therefore,

$$f(\mathbf{x}_0,y)\le w(\mathbf{x}_0,y)+k_2(y)\le \frac{2M}{H}+k_2(y)<2\epsilon+k_\omega(y)$$

or $f(\mathbf{x}_0,y)-k(y,\omega)<2\epsilon$.

Together with a similar argument using lower barriers and $k_1(y)$ (i.e., $u_1(\mathbf{x},y)=l_a(\mathbf{x},y)+k_1(y)$ with $\Psi(\rho)=1$), we then find that

$$|f(\mathbf{x}_0,y)-k(y,\omega)|<2\epsilon.$$

Since $\mathbf{x}_0\in W$ is arbitrary, we finally have

$$|f(\mathbf{x},y)-k(y,\omega)|\le 2\epsilon \quad\text{for}\quad (\mathbf{x},y)\in\Omega \quad\text{with}\quad \mathbf{x}\in W. \tag{32}$$

Now if $\frac{\mathbf{x}_j}{|\mathbf{x}_j|}\to\omega$ as $j\to\infty$, there exists $N>0$ such that $\mathbf{x}_j\in W$. Then from (32), for $(\mathbf{x}_j,y_j)\in\Omega$, we have

$$|f(\mathbf{x}_j,y_j)-k(y_j,\omega)|\le 2\epsilon \quad\text{if}\quad j\ge N.$$

Since $\epsilon>0$ is arbitrary, the conclusion of Theorem 2.2 follows.

Proof of Theorem 2.3. Consider first the following:

Lemma 4.1. *Suppose $M>0$, $I=(a,b)\subset I_M$, $E=E(y,z,q)$ is a nonincreasing function of z for each $(y,q)\in\overline{I}\times\mathbf{R}$, and E, $\frac{\partial E}{\partial z}$, $\frac{\partial E}{\partial q}\in C^0(\overline{I}\times\mathbf{R}^2)$. Suppose also that there exists $k\in C^2(\overline{I})$ which satisfies*

$$k''(y)+E(y,k(y),k'(y))=0 \qquad \textit{for}\quad y\in I.$$

Then for each $\delta_1>0$, there is a number $\beta>0$ such that if $c\in\mathbf{R}$ with $|c|<\beta$, then there exists $k_{(c)}\in C^2(\overline{I})$ satisfying

$$k''_{(c)}(y)+E(y,k_{(c)}(y),k'_{(c)}(y))=c,\quad k_{(c)}(a)=k(a),\quad k_{(c)}(b)=k(b),$$

and

$$|k(y)-k_{(c)}(y)|\le\delta_1 \quad \textit{for}\quad y\in I.$$

Using Lemma 4.1, whose proof is given in the appendix, we see that the hypotheses of Theorem 2.2 are satisfied and then Theorem 2.3 is proven.

5. Examples.

There are many common examples of operators of the form (1), normalized to satisfy (5), which satisfy (14). Some of these are the (normalized a la (5)) Laplace, Poisson, minimal surface, prescribed mean curvature, p-Laplace (for C^2 solutions), and heat (e.g., with $t = x_1$) operators. A C^2 solution of a fully nonlinear equation may also be considered here when the appropiate (normalized) quasilinear operator (i.e., [**4**] (17.10)) satisfies the hypotheses of our Theorems (i.e., p. 444, [**4**]).

Example 5.1. Suppose $n = 2$, $\Omega = \{(x, y) : x > 0, -M < y < M\}$ for some $M > 0$, $h \in C^0(\overline{\Omega} \times \mathbf{R}^3)$, $h(x, y, z, p, q) = m(y) + o(1)$ as $x \to \infty$ uniformly for $|y| \le M$ and $z, p, q \in \mathbf{R}$, $\max_{y\in\overline{I_M}} |\int_0^y m(s)ds| = \alpha_0 < 1$, $\phi(x, \pm M) \to 0$ as $x \to \infty$, and Q is a mean curvature operator with $Qu(x, y)$ equal to

$$\frac{(1 + u_y^2)u_{xx} - 2u_x u_y u_{xy} + (1 + u_x^2)u_{yy}}{2 + u_x^2 + u_y^2} - \frac{h(x, y, u, u_x, u_y)(1 + u_x^2 + u_y^2)^{\frac{3}{2}}}{2 + u_x^2 + u_y^2}.$$

When f is a solution of (2) & (3), ϕ is bounded, and

$$2M|h(x, y, f(x, y), f_x(x, y), f_y(x, y))| \le \beta_0 < 1$$

for all $(x, y) \in \Omega$, the use of comparison arguments with Delaunay surfaces shows that f is bounded. Notice that $a_{22}(x, y, z, p, q) = \frac{1+p^2}{2+p^2+q^2}$ and $E(\omega, y, q) = -m(y)\big(1 + q^2\big)^{\frac{3}{2}}$. The Dirichlet problem for (13) is

(33)

$$k''(y) = m(y)\big(1 + (k'(y))^2\big)^{\frac{3}{2}} \text{ for } |y| \le M \text{ with } k(-M) = k(M) = 0.$$

Suppose $m(y) > 0$ for $y \in \mathbf{R}$ and $\int_{-M}^{M} m(y)\, dy < 1$. Theorem 2.4 of [**18**] implies (33) has a unique classical solution $k(y)$ and Theorem 2.3 implies

$$\lim_{x\to\infty} f(x, y) = k(y) \tag{34}$$

when $|y| \le M$ for any bounded solution f of (2) & (3). On the other hand, if we had $m(y) > 0$ for $y \in \mathbf{R}$ and $\int_{-M}^{M} m(y)\, dy \ge 1$, Theorem 2.3 (ii) of [**18**] would imply (33) has no solution in $C^1(\overline{I_M}) \cap C^2(I_M)$.

Example 5.2. Suppose $n = 3$, $M = 1$, $\Omega = S_M$, Q is defined by

$$Qu(x_1, x_2, y) = \frac{1}{3}\big(u_{x_1x_1} + u_{x_2x_2} + u_{yy}\big) - \frac{1}{3}\left(u_y^2 + \frac{x_1^2}{1 + |\mathbf{x}|^2}\right),$$

and $\phi(\mathbf{x}, \pm 1) = 0$ for $|\mathbf{x}| \ge 1$. Notice that $E(\omega, y, z, q) = -(q^2 + \omega_1^2)$ for $\omega = (\omega_1, \omega_2) \in S^1$. The Dirichlet problem (13) here is $k''_\omega(y) = (k'_\omega(y))^2 + \omega_1^2$ with $k_\omega(-1) = k_\omega(1) = 0$ and its solution is

$$k(y, \omega) = \ln\big(\sec(\omega_1 y)\big) - \ln\big(\sec(\omega_1)\big).$$

If $f \in C^2(\Omega) \cap C^0(\overline{\Omega})$ is a bounded solution of (2) & (3), Theorem 2.3 implies, for example, that

$$\lim_{r\to\infty} f(r\omega, y) = k(y,\omega) = \ln\big(\sec(\omega_1 y)\big) - \ln\big(\sec(\omega_1)\big)$$

for $|y| \le 1$.

In the next example, the domain has the form $\Omega = U \times I_M$ with U a subset of $\mathbf{R}^2$ which contains the first quadrant of the plane and whose boundary oscillates sinusoidally in the second and fourth quadrants. The conclusion of Theorem 2.2 is satisfied for directions into the (open) first quadrant but not for directions into the (closed) second or fourth quadrant. The behavior at infinity of a bounded solution in an "oscillatory direction" ω for such a domain is an open question.

Example 5.3. Let $n = 3$, $M = 1$, and set $h(r) = \frac{\pi}{4}(1 + \sin(r))$. Let

$$\Omega = \left\{(r\cos(\theta), r\sin(\theta), y) \in \mathbf{R}^3 : r > 0, -h(r) < \theta < \frac{\pi}{2} + h(r), |y| < 1\right\}.$$

Notice that the set of directions at infinity for Ω is $T = \{(\cos(\theta), \sin(\theta)) : \theta \in [-\frac{\pi}{2}, \pi]\}$. Set $T_0 = \{(\cos(\theta), \sin(\theta)) : \theta \in (0, \frac{\pi}{2})\}$ and $T_1 = T \setminus T_0$. Define Q by $Qu = \frac{1}{3}(\triangle u - u)$ and $\phi \equiv \cosh(1)$; notice that $E(\omega, y, z, q) = -z$. Let $f \in C^2(\Omega) \cap C^0(\overline{\Omega})$ denote a bounded solution of (2) & (3).

Suppose first that $\omega \in T_0$. Then (8) and (13) yield

$$k_\omega''(y) - k_\omega(y) = 0 \quad \text{for} \quad -1 < y < 1, \qquad k_\omega(\pm 1) = \cosh(1),$$

and so $k_\omega(y) = \cosh(y)$. Setting $k_1(y) = \cosh((1+\epsilon)y)$ and $k_2(y) = \cosh((1-\epsilon)y)$ for $\epsilon > 0$ sufficiently small shows that Assumption 2 is satisfied. Theorem 2.2 then implies

$$f(\mathbf{x}, y) \to \cosh(y)$$

as $|\mathbf{x}| \to \infty$ with $\frac{\mathbf{x}}{|\mathbf{x}|} \to \omega$ uniformly for $|y| \le 1$.

Suppose second that $\omega \in T_1$. Notice that (8) requires $k_\omega(y) = \cosh(1)$ for all $y \in [-1, 1]$. However this constant function is not a solution of (13). In fact, it is impossible to obtain a function $k_1(y)$ which satisfies (9) and (11) when α is sufficiently small. (Notice that (9) implies $k_1(\pm 1) \le \cosh(1) + \alpha$ and (11) then implies $k_1(y) \le \left(1 + \frac{\alpha}{\cosh(1)}\right)\cosh(y)$. On the other hand, (9) implies $k_1(y) \ge \cosh(1) - \alpha$ and so $\cosh(1) - \alpha \le k_1(0) \le 1 + \frac{\alpha}{\cosh(1)}$; this is impossible for $\alpha > 0$ sufficiently small.) This means that the hypotheses of Theorem 2.2 are not satisfied when $\omega \in T_1$. (If the Dirichlet data had satisfied $\phi(\mathbf{x}, y) \to \cosh(y)$ as $\mathbf{x} \to \infty$, then Theorem 2.2 would have been applicable for all directions $\omega \in T$ and our conclusion would be that $f(\mathbf{x}, y) \to \cosh(y)$ as $|\mathbf{x}| \to \infty$ uniformly for $|y| \le 1$.) If we set $k_1(y) = \cosh(y)$ and $k_2(y) = \cosh(1)$, the comparison argument in the Proof of Theorem 2.2 shows that for $|y| \le 1$ and $\omega \in T_1$,

$$\cosh(y) \le \liminf_{r\to\infty} f(r\omega, y) \le \limsup_{r\to\infty} f(r\omega, y) \le \cosh(1).$$

A general characterization of the behavior of a bounded solution of (2) & (3) when the boundary oscillates in a manner similar to that considered here would be very interesting.

Using our techniques, some structural conditions on Q which imply that all solutions $f \in C^2(\Omega) \cap C^0(\overline{\Omega})$ of (2) & (3) are bounded can be obtained. Although we lmit our discussion here primarily to domains in slabs S_M, the geometric condition that $\Omega \subset S_M$ can be weakened substantially (e.g., §5, [**14**]) without changing the conclusion that $f(\mathbf{x}, y) \to k_\omega(y)$ as $|\mathbf{x}| \to \infty$. Theorem 2.2 can also be applied to determine the asymptotic behavior of solutions of Dirichlet problems in exterior domains for certain types of operators.

Appendix.

Proof of Lemma 4.1. We may assume $I = I_M$. Suppose first that a function $k_c(y)$ satisfying the conclusion of Lemma 4.1 did exist. If we were to set $s(y) = k_c(y) - k(y)$, then $s(y)$ would satisfy $|s(y)| \le \delta_1$ for $|y| \le M$ and

$$s''(y) + k''(y) + E(y, k(y) + s(y), k'(y) + s'(y)) = c \qquad \text{with } \ s(-M) = 0, \quad s(M) = 0.$$

If we define $G(y, z, q) = E(y, k(y)+z, k'(y)+q) - E(y, k(y), k'(y))$ and recall that $k''(y) + E(y, k(y), k'(y)) = 0$, we see that $s(y)$ would satisfy

$$s(-M) = 0, \qquad s(M) = 0, \tag{35}$$

$$s''(y) + G(y, s(y), s'(y)) = c, \tag{36}$$

and

$$|s(y)| \le \delta_1 \quad \text{for} \quad |y| \le M. \tag{37}$$

Conversely, if we find a function $s \in C^2(\overline{I})$ which satisfies (36), (37), $s(-M) = 0$, and $s(M) = 0$, then the function $k_c(y) = k(y) + s(y)$ satisfies the conclusion of Lemma 4.1.

Let us rewrite (36) as

$$\begin{aligned} s''(y) &+ \frac{\partial G}{\partial q}(y,0,0)s'(y) + \frac{\partial G}{\partial z}(y,0,0)s(y) \\ &= c + \frac{\partial G}{\partial q}(y,0,0)s'(y) + \frac{\partial G}{\partial z}(y,0,0)s(y) - G(y, s(y), s'(y)). \end{aligned}$$

We define a sequence $\{s_n\}$ by

$$s_1(y) = 0 \quad \textit{on} \quad |y| \le M$$

and, for $n \geq 1$,

$$s''_{n+1}(y) + \frac{\partial G}{\partial q}(y,0,0)s'_{n+1}(y) + \frac{\partial G}{\partial z}(y,0,0)s_{n+1}(y) \tag{38}$$
$$= c + \frac{\partial G}{\partial q}(y,0,0)s'_n(y) + \frac{\partial G}{\partial z}(y,0,0)s_n(y) - G(y, s_n(y), s'_n(y))$$

with

$$s_{n+1}(-M) = 0, \quad s_{n+1}(M) = 0. \tag{39}$$

We claim that when $|c|$ is small enough, the sequence $s_n(y)$ converges uniformly on $[-M, M]$ to a function $s \in C^2(\bar{I})$ which satisfies (36), (37), and (35).

We require several estimates. Consider the boundary value problem

$$w''(y) + \frac{\partial G}{\partial q}(y,0,0)w'(y) + \frac{\partial G}{\partial z}(y,0,0)w(y) = h(y),$$
$$w(-M) = 0, \quad w(M) = 0.$$

Since $E(y,z,q)$ is non-increasing on z, $\frac{\partial G}{\partial z}(y,0,0) = \frac{\partial E}{\partial z}(y,k(y),k'(y)) \leq 0$ for all $|y| \leq M$. Now we apply Theorem 3.7 (or the proof of Theorem 3.7) in [**4**] to conclude that there is a constant c_1 depending on E and $k(y)$ such that

$$|w(y)| \leq c_1 |h|_{C^0([-M,M])} \quad on \quad |y| \leq M; \tag{40}$$

for notational simplicity, we will write $\|u\|$ or $\|u(y)\|$ for the value of the supremum norm $|u|_{C^0([-M,M])}$ of a function u. Using the equation

$$w''(y) + \frac{\partial G}{\partial q}(y,0,0)w'(y) = h(y) - \frac{\partial G}{\partial z}(y,0,0)w(y), \tag{41}$$

we see that

$$w'(y) = \int_{-M}^{y} \exp\left(\int_y^t \frac{\partial G}{\partial q}(\alpha,0,0)d\alpha\right)\left(h(t) - \frac{\partial G}{\partial z}(t,0,0)w(t)\right)dt$$
$$+ B\exp\left(-\int_{-M}^{y} \frac{\partial G}{\partial q}(\alpha,0,0)d\alpha\right)$$

where

$$B = -\frac{\int_{-M}^{M}\int_{-M}^{y} \exp\left(\int_y^t \frac{\partial G}{\partial q}(\alpha,0,0)d\alpha\right)\left(h(t) - \frac{\partial G}{\partial z}(t,0,0)w(t)\right)dt\ dy}{\int_{-M}^{M} \exp\left(-\int_{-M}^{y} \frac{\partial G}{\partial q}(\alpha,0,0)d\alpha\right)dy}.$$

Using (40), we see that for some constant c_2,

$$|w'(y)| \leq c_2 \|h\| \quad \text{for} \quad |y| \leq M. \tag{42}$$

Using (40), (42), and

$$w''(y) = h(y) - \frac{\partial G}{\partial q}(y,0,0)w'(y) - \frac{\partial G}{\partial z}(y,0,0)w(y),$$

we conclude that there is a constant c_3 depending only on M, G (hence E, k) such that

$$|w(y)| \le c_3\|h\| \quad \text{for} \quad |y| \le M; \tag{43}$$

$$|w'(y)| \le c_3\|h\| \quad \text{for} \quad |y| \le M; \tag{44}$$

$$|w''(y)| \le c_3\|h\| \quad \text{for} \quad |y| \le M. \tag{45}$$

Setting $w(y) = s_{n+1}(y)$ in (43)-(45), for $|y| \le M$ we obtain

$$\begin{aligned}
|s_{n+1}(y)| \le c_3\Big(|c| + \Big\| &\frac{\partial G}{\partial q}(y,0,0)s_n'(y) + \frac{\partial G}{\partial z}(y,0,0)s_n(y) \\
&- G(y, s_n(y), s_n'(y))\Big\|\Big) \\
|s_{n+1}'(y)| \le c_3\Big(|c| + \Big\| &\frac{\partial G}{\partial q}(y,0,0)s_n'(y) + \frac{\partial G}{\partial z}(y,0,0)s_n(y) \\
&- G(y, s_n(y), s_n'(y))\Big\|\Big) \\
|s_{n+1}''(y)| \le c_3\Big(|c| + \Big\| &\frac{\partial G}{\partial q}(y,0,0)s_n'(y) + \frac{\partial G}{\partial z}(y,0,0)s_n(y) \\
&- G(y, s_n(y), s_n'(y))\Big\|\Big).
\end{aligned}$$

Now since $\frac{\partial E}{\partial q}(y,z,q)$ and $\frac{\partial E}{\partial z}(y,z,q)$ are continuous on $[-M,M] \times R^2$, $\frac{\partial E}{\partial q}(y,z,q)$ and $\frac{\partial E}{\partial z}(y,z,q)$ are uniformly continuous on $[-M,M] \times D$ for any compact set $D \subset R^2$. Since $k(y) \in C^2([-M,M])$, it follows that $\frac{\partial G}{\partial q}(y,z,q)$ and $\frac{\partial G}{\partial z}(y,z,q)$ are uniformly continuous on $[-M,M] \times [-1,1] \times [-1,1]$. Then there is a constant $\gamma > 0$ such that

$$\left|\frac{\partial G}{\partial q}(y,0,0) - \frac{\partial G}{\partial q}(y,\alpha_1,\alpha_2)\right| \le \frac{1}{8c_3} \tag{46}$$

$$\left|\frac{\partial G}{\partial z}(y,0,0) - \frac{\partial G}{\partial z}(y,\alpha_1,\alpha_2)\right| \le \frac{1}{8c_3} \tag{47}$$

for all $|\alpha_1| \le \gamma$, $|\alpha_2| \le \gamma$, $|y| \le M$. Now set

$$\delta = \min\{\delta_1, \gamma\}, \quad \delta_2 = \frac{3}{4c_3}\delta.$$

We claim that for $|c| \le \delta_2$,

$$|s_n(y)| \le \delta, \quad |s_n'(y)| \le \delta, \quad |s_n''(y)| \le \delta \tag{48}$$

for all n, $|y| \le M$. Indeed, if $n = 1$, (48) is obvious since $s_1(y) = 0$. Now we assume (48) holds for all integers up to n. In order to prove (48) for $n+1$, we need only show that

$$c_3\left(|c| + \left\|\frac{\partial G}{\partial q}(y,0,0)s_n'(y) + \frac{\partial G}{\partial z}(y,0,0)s_n(y) - G(y, s_n(y), s_n'(y))\right\|\right) \le \delta.$$

Since $G(y,0,0) = 0$, we have from the mean value theorem

$$\begin{aligned}
&G(y, s_n(y), s_n'(y))\\
&= G(y, s_n(y), s_n'(y)) - G(y,0,0)\\
&= \frac{\partial G}{\partial q}(y, \theta(y)s_n(y), \theta(y)s_n'(y))s_n'(y) + \frac{\partial G}{\partial z}(y, \theta(y)s_n(y), \theta(y)s_n'(y))s_n(y)
\end{aligned}$$

for some function $\theta(y)$ on $[-M, M]$ with $|\theta(y)| \le 1$. Then from (46), (47) and (48), we have

$$\begin{aligned}
&\left\|\frac{\partial G}{\partial q}(y,0,0)s_n'(y) + \frac{\partial G}{\partial z}(y,0,0)s_n(y) - G(y, s_n(y), s_n'(y))\right\|\\
&\le \left\|\frac{\partial G}{\partial q}(y,0,0) - \frac{\partial G}{\partial q}(y, \theta(y)s_n(y), \theta(y)s_n'(y))\right\| \cdot \|s_n'\|\\
&\quad + \left\|\frac{\partial G}{\partial z}(y,0,0) - \frac{\partial G}{\partial z}(y, \theta(y)s_n(y), \theta(y)s_n'(y))\right\| \cdot \|s_n\| \le \frac{1}{4c_3}\delta.
\end{aligned}$$

Hence

$$\begin{aligned}
&c_3\left(|c| + \left\|\frac{\partial G}{\partial q}(y,0,0)s_n'(y) + \frac{\partial G}{\partial z}(y,0,0)s_n'(y) - G(y, s_n(y), s_n'(y))\right\|\right)\\
&\le c_3\left(|c| + \frac{1}{4c_3}\delta\right) \le c_3\left(\delta_2 + \frac{1}{4c_3}\delta\right) = \delta.
\end{aligned}$$

This implies (48) holds for all n.

We claim that the sequence $s_n(y)$ converges uniformly on $[-M, M]$. Let us set $w(y) = s_{n+1}(y) - s_n(y)$. From (38), we have

$$\begin{aligned}
w''(y) + \frac{\partial G}{\partial q}(y,0,0)w'(y) + \frac{\partial G}{\partial z}(y,0,0)w(y) = h(y),\\
w(-M) = 0, \quad w(M) = 0,
\end{aligned}$$

where the function h is defined by

$$\begin{aligned}
h(y) = &\frac{\partial G}{\partial q}(y,0,0)(s_n(y) - s_{n-1}(y))' + \frac{\partial G}{\partial z}(y,0,0)(s_n(y) - s_{n-1}(y))\\
&- (G(y, s_n(y), s_n'(y)) - G(y, s_{n-1}(y), s_{n-1}'(y))).
\end{aligned}$$

Thus from (43)-(45), we have the inequalities

$$
\begin{aligned}
&|s_{n+1}(y)-s_n(y)| \quad \text{and} \quad |s'_{n+1}(y)-s'_n(y)| \quad \text{and} \quad |s''_{n+1}(y)-s''_n(y)| \\
&\le c_3 \bigg| \frac{\partial G}{\partial q}(y,0,0)(s_n(y)-s_{n-1}(y))' + \frac{\partial G}{\partial z}(y,0,0)(s_n(y)-s_{n-1}(y)) \\
&\quad -(G(y,s_n(y),s'_n(y)) - G(y,s_{n-1}(y),s'_{n-1}(y)))\bigg|_{C^0([-M,M])}.
\end{aligned}
$$

Setting $r(y) = s_n(y) - s_{n-1}(y)$ for a moment and using the mean value theorem, we see that the sum above is bounded by

$$
\begin{aligned}
&c_3 \bigg[\bigg\| \frac{\partial G}{\partial q}(y,0,0) - \frac{\partial G}{\partial q}(y, s_n(y)+\eta(y)r(y), s'_n(y)+\eta(y)r'(y)) \bigg\| \; \|r'(y)\| \\
&+ \bigg\| \frac{\partial G}{\partial z}(y,0,0) - \frac{\partial G}{\partial z}(y, s_n(y)+\eta(y)r(y), s'_n(y)+\eta(y)r'(y)) \bigg\| \; \|r(y)\| \bigg],
\end{aligned}
$$

for some functions $\eta(y)$ on $[-M,M]$ with $|\eta(y)| \le 1$. From (46) and (47), we see that this sum is bounded by

$$
\frac{1}{8}\|s'_n(y) - s'_{n-1}(y)\| + \frac{1}{8}\|s_n(y) - s_{n-1}(y)\|;
$$

hence for $|y| \le M$,

$$
\begin{aligned}
|s_{n+1}(y)-s_n(y)| &\le \frac{1}{8}\|s'_n - s'_{n-1}\| + \frac{1}{8}\|s_n - s_{n-1}\|, \\
|s'_{n+1}(y)-s'_n(y)| &\le \frac{1}{8}\|s'_n - s'_{n-1}\| + \frac{1}{8}\|s_n - s_{n-1}\|, \qquad \text{and} \\
|s''_{n+1}(y)-s''_n(y)| &\le \frac{1}{8}\|s'_n - s'_{n-1}\| + \frac{1}{8}\|s_n - s_{n-1}\|.
\end{aligned}
$$

Setting

$$
C = \|s_2(y) - s_1(y)\| + \|s'_2(y) - s'_1(y)\| + \|s''_2(y) - s''_1(y)\|
$$

and using the inequalities above, we observe that an induction argument yields

$$
\begin{aligned}
|s_{n+1} - s_n|_{C^0([-M,M])} &\le C\left(\frac{1}{4}\right)^{n-1}, \\
|s'_{n+1} - s'_n|_{C^0([-M,M])} &\le C\left(\frac{1}{4}\right)^{n-1}, \\
|s''_{n+1} - s''_n|_{C^0([-M,M])} &\le C\left(\frac{1}{4}\right)^{n-1},
\end{aligned}
$$

for all $n \ge 2$.

Therefore $s_n(y)$, $s'_n(y)$ and $s''_n(y)$ converge uniformly on $[-M,M]$. Let

$$
s(y) = \lim_{n \longrightarrow \infty} s_n(y).
$$

Then it is easy to see $s(y)$ is in $C^2([-M, M])$ and satisfies (36)-(37). This completes the proof.

Acknowledgement. The authors would like to thank Robert Finn and Ronald Guenther for their interest in and comments on preliminary versions of this paper.

References

[1] H. Bear and G. Hile, *Behavior of solutions of elliptic differential inequalities near a point of discontinuous boundary data*, Comm. Part. Diff. Equ., **8** (1983), 1175-1197, MR 85g:35054, Zbl 0533.35040.

[2] S. Breuer and J. Roseman, *Decay theorems for nonlinear Dirichlet problems in semi-infinite cylinders*, Arch. Rat. Mech. Anal., **94** (1986), 363-371, MR 88g:35031, Zbl 0613.35029.

[3] L. Gårding, *Some Points of Analysis and Their History*, University Lecture Series, **11**, Amer. Math. Soc., Providence, RI, 1997, MR 98m:01023, Zbl 0885.35001.

[4] D. Gilbarg and N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, second edition, Springer-Verlag, 1983, MR 86c:35035, Zbl 0562.35001.

[5] A. Granas, R. Guenther and J. Lee, *Nonlinear Boundary Value Problems for Ordinary Differential Equations*, Dissertationes Mathematicae, **CCXLIV**, 1985, MR 87a:34025, Zbl 0615.34010.

[6] ______, *Some general existence principles in the Carateheódory theory of nonlinear differential systems*, J. Math. Pures et Appl., **70** (1991), 153-196, MR 92d:34041, Zbl 0687.34009.

[7] C. Horgan and J. Knowles, *The effect of nonlinearity on a principle of Saint-Venent type*, J. Elasticity, **11** (1981), 271-291, MR 82h:73041, Zbl 0463.73007.

[8] C. Horgan and L. Payne, *Decay estimates for a class of second-order quasilinear equations in three dimensions*, Arch. Rat. Mech. Anal., **86** (1984), 279-289, MR 86d:35018, Zbl 0567.35016.

[9] ______, *On the asymptotic behavior of solutions of inhomogeneous second-order quasilinear partial differential equations*, Quart. Appl. Math., **47** (1989), 753-771, MR 90m:35033, Zbl 0696.35011.

[10] ______, *The effect of constitutive law perturbations on finite antiplane shear deformations of a semi-infinite strip*, Quart. Appl. Math., **51** (1993), 441-465, MR 94f:73022, Zbl 0781.73016.

[11] ______, *Spatial decay estimates for a class of second-order quasilinear elliptic partial differential equations arising in anisotropic nonlinear elasticity*, Math. Mech. Solids, **1** (1996), 411-423, MR 97i:73030, Zbl 1001.74524.

[12] J. Hwang, *Phragmén-Lindelöf theorem for the minimal surface equation*, Proc. Amer. Math. Soc., **104** (1988), 825-828, MR 89j:35016, Zbl 0787.35012.

[13] Z. Jin and K. Lancaster, *Theorems of Phragmén-Lindelöf type for quasilinear elliptic equations*, J. Reine Angew. Math., **514** (1999), 165-197, MR 2001c:35074, Zbl 0937.35053.

[14] ———, *Phragmén-Lindelöf Theorems and the Asymptotic Behavior of solutions of quasilinear elliptic equations in slabs*, Proc. Roy. Soc. Edinburgh, **130A** (2000), 335-373, MR 2001f:35128, Zbl 0965.35069.

[15] ———, *The convergence rate of solutions of quasilinear elliptic equations in slabs*, Comm. Part. Diff. Equ., **25** (2000), 957-985, MR 2001b:35110, Zbl 0947.35057.

[16] J. Knowles, *A note on the spatial decay of a minimal surface over a semi-infinite strip*, J. Math. Anal. Appl., **59** (1977), 29-32, MR 55 #10842, Zbl 0356.35032.

[17] J.C.C. Nitsche, *On new results in the theory of minimal surfaces*, Bull. Amer. Math. Soc., **71** (1965), 195-270, MR 30 #4200, Zbl 0135.21701.

[18] A. O'Farrell and D. O'Regan, *Existence results for some initial- and boundary-value problems*, Proc. Amer. Math. Soc., **110** (1990), 661-673, MR 91b:34038, Zbl 0713.34024.

[19] D. O'Regan, *Existence Theory for Nonlinear Ordinary Differential Equations*, Mathematics and Its Applications, **398**, Kluwer, 1997, MR 98h:34042.

[20] E. Phragmén and E. Lindelöf, *Sur une extension d'un principe classique de l'analyse et sur quelque propriétès des fonctions monogènes dans le voisinage d'un point singulier*, Acta Mathematica, **31** (1908), 381-406.

[21] J. Roseman, *Phragmén-Lindelöf theorems for some non-linear elliptic partial differential equations*, J. Math. Anal. Appl., **43** (1973), 587-602, MR 49 #804, Zbl 0281.35003.

[22] J. Serrin, *The problem of Dirichlet for quasilinear equations with many independent variables*, Phil. Trans. Royal Soc. London Ser. A, **264** (1969), 413-496, MR 43 #7772, Zbl 0181.38003.

Received March 18, 2002.

Department of Mathematics and Statistics
Wichita State University
Wichita, KS 67260-0033
E-mail address: zhiren@math.twsu.edu

Department of Mathematics and Statistics
Wichita State University
Wichita, KS 67260-0033
E-mail address: lancaster@math.twsu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

ON PRIMITIVE SUBDIVISIONS OF AN ELEMENTARY TETRAHEDRON

J.-M. Kantor and K.S. Sarkaria

A polytope P of 3-space, which meets a given lattice $\mathbb{L}$ only in its vertices, is called $\mathbb{L}$-elementary. An $\mathbb{L}$-elementary tetrahedron has volume $\geq (1/6).\det(\mathbb{L})$, in case equality holds it is called $\mathbb{L}$-primitive. A result of Knudsen, Mumford and Waterman, tells us that any convex polytope P admits a linear simplicial subdivision into tetrahedra which are primitive with respect to the bigger lattice $(1/2)^t.\mathbb{L}$, for some t depending on P. Improving this, we show that in fact the lattice $(1/4).\mathbb{L}$ always suffices. To this end, we first characterize all $\mathbb{L}$-elementary tetrahedra for which even the intermediate lattice $(1/2).\mathbb{L}$ suffices.

1. Introduction.

Following Danilov [**Da**], a polytope will be called ELEMENTARY with respect to a given lattice $\mathbb{L}$ (or $\mathbb{L}$-elementary, or just elementary) iff it meets the lattice $\mathbb{L}$ only in its vertices. For example, it is easy to see that a parallelepiped with vertices in a lattice $\mathbb{L} \subset \mathbb{R}^3$ is elementary with respect to $\mathbb{L}$ if and only if its volume equals $|\det \mathbb{L}|$.

A tetrahedron will be called PRIMITIVE[1] with respect to a lattice $\mathbb{L}$ (or $\mathbb{L}$-primitive, or just primitive) iff its vertices are in $\mathbb{L}$ and its volume is $\frac{1}{6}|\det \mathbb{L}|$. *A primitive tetrahedron is elementary* because it is one of the 6 tetrahedra of a volume $|\det \mathbb{L}|$ parallelepiped with vertices in $\mathbb{L}$. *The converse is false.* One obtains a $\mathbb{Z}^3$-elementary tetrahedron ABCD whenever a line segment AB, lying in the plane $z = 0$, and having no integral points except A and B, is joined to an analogous general position line segment CD lying in the "adjacent" parallel plane $z = 1$, and such tetrahedra can have volume bigger than any given number.

[1]Usage varies a lot, e.g., primitive is "elementary" in [**GKZ**], and "basic" and "minimal" have also been used; likewise, elementary is "fundamental" in [**Re**, **Rez**], "admissible" in [**Wh**], "lattice free" in [**Ka2**], etc.

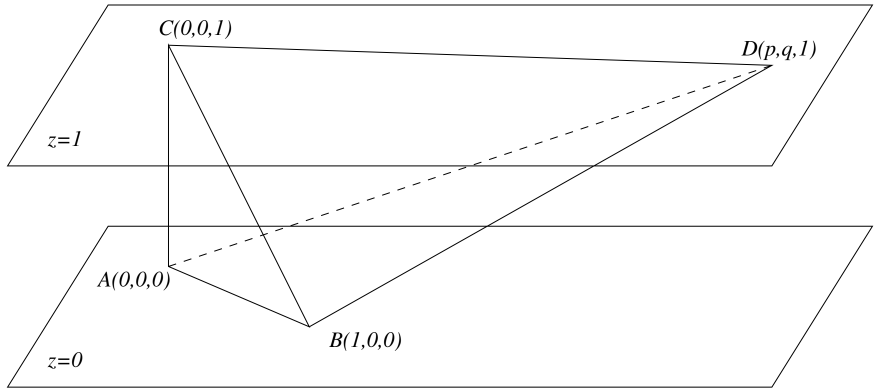


Figure 1.

A theorem of White [**Wh**] assures us that, upto an ISOMORPHISM, i.e., upto an affine linear transformation which maps $\mathbb{Z}^3$ onto itself, the above construction yields all $\mathbb{Z}^3$-elementary tetrahedra. It follows (by an argument similar to the proof of Lemma 2 of §2) that, with respect to a suitable origin and basis, we can assume the four vertices to be $A = (0,0,0)$, $B = (1,0,0)$, $C = (0,0,1)$ and $D = (p,q,1)$, where $0 \leq p \leq q$ with g.c.d. $(p,q) = 1$. These standard models of elementary tetrahedra — Figure 1 — will be denoted $\mathbf{T}(p,q)$. Note that the tetrahedron $\mathbf{T}(p,q)$ has volume $q/6$; so, for all $q \geq 2$, it is non-primitive; the primitive case corresponds to $p = 0, 1$ and $q = 1$, otherwise $1 \leq p < q$.

We note here that in the two dimensional case, the analogues of the above two notions would coincide: *An $\mathbb{L}$-elementary triangle must have area* $\frac{1}{2}|\det \mathbb{L}|$. For, by adjoining an isomorphic triangle, one can make a parallelogram which is elementary, and so has area $|\det \mathbb{L}|$.

In some arithmetical questions one uses convex cell complexes in $\mathbb{R}^n$ having integral vertices. These admit an ELEMENTARY SUBDIVISION with respect to the lattice $\mathbb{Z}^n$, i.e., can be linearly subdivided, e.g., by using the method of Hudson [**Hu**], p. 11, into simplicial complexes having only $\mathbb{Z}^n$ -elementary simplices. However, for $n \geq 3$, as above examples of non-primitive elementary tetrahedra already show, not always into primitive simplices. So, to obtain a PRIMITIVE SUBDIVISION, it is necessary to consider, instead of $\mathbb{Z}^n$, a bigger lattice, say that of all half-integral points, for which we have the following result:

Theorem 1. *A* $\mathbf{T}(p,q)$, $q \neq 1$, *admits a linear simplicial subdivision, primitive with respect to* $\frac{1}{2}\mathbb{Z}^3$, *iff* $p = 1$ *or* $p = q - 1$.

It is easily seen that $\mathbf{T}(1,q)$ and $\mathbf{T}(q-1,q)$ are isomorphic, so this theorem tells us that, for each $q \geq 2$, there is, upto isomorphism, just one non-primitive elementary tetrahedron which can be primitively subdivided by using half-integral points. The situation changes dramatically when we are allowed to use quarter-integral points.

Theorem 2. *Any* $\mathbf{T}(p,q)$ *admits a linear simplicial subdivision, primitive with respect to the lattice* $\frac{1}{4}\mathbb{Z}^3$.

Moreover, our method allows, as one passes in the subdivision process from $\mathbb{Z}^3$ to $\frac{1}{2}\mathbb{Z}^3$ and then from $\frac{1}{2}\mathbb{Z}^3$ to $\frac{1}{4}\mathbb{Z}^3$, to subdivide the faces of the tetrahedron in a *standard way* (each triangular face is cut up into four smaller triangles by joining the mid-points of its three sides to each other). This stronger **Theorem 2′** ensures that the subdivisions of the constituent elementary simplices of a complex fit together, i.e., that one also has the following more general result:

Theorem 3. *A convex cell complex with vertices in* $\mathbb{Z}^3$ *always admits a* $\frac{1}{4}\mathbb{Z}^3$*-primitive subdivision.*

This improves on a theorem of Knudsen, Mumford and Waterman — given in Kempf-Knudsen-Mumford-St.Donat [**KKMS**] — which asserts the existence of a $(\frac{1}{2})^t\mathbb{Z}^3$-primitive subdivision for some integer t depending on the given cell complex. However, we note that these authors prove their result for any $\mathbb{R}^n$, and that they check that their linear subdivisions are projective,[2] an aspect of the matter which we will ignore in this paper; also (see §3) this weaker result generalizes to polyhedral non-convex cells, while convexity of the cells is needed if one wants to use only quarter-integral points.

Note by J.-M. Kantor. The method of this paper was worked out completely, without any knowledge of previous work, by K.S. Sarkaria. It was only after this paper was written up (as preprint IHES/M/01/23) that I remembered, and passed on to my co-author, some quite incomprehensible drawings which had been given to me by Ziegler, and which apparently had been shown during a talk given by him in 1997.

Note by K.S. Sarkaria. The problem settled by Theorem 3 was first posed to me in 1995 by J.-M. Kantor; however what eventually sparked this work (done during the time period November 2000 to January 2001) were some stimulating conversations which I had with him much later, during the month of October 2000, when I was visiting Bures-sur-Yvette. I am also grateful to Jean-Michel for checking the arguments of this paper.

Subsequently, I have learnt, by deciphering the aforementioned "incomprehensible drawings", that this work connects with unpublished (actually still to be written out) work of Jeff Lagarias and Günter Ziegler done in

[2]We refer the reader to [**KKMS**] for this definition, as well as for the importance of these results for the theory of toric varieties.

1997, e.g., they had a different method, which apparently also gives a $\mathbb{Z}^3$-primitively subdivided $4\mathbf{T}(p,q)$, however *with boundary not subdivided in a standard way,* thus it does not yield Theorem 3. I suspect that, if one does not impose any condition on the boundary subdivision, then even $3\mathbf{T}(p,q)$ is $\mathbb{Z}^3$-primitively triangulable. In any case, I have checked this for $q \equiv \pm 1 \bmod p$. All in all, dispensing with the boundary condition should make many tetrahedral problems much easier, a theme I hope to pursue elsewhere.

Outline of the paper. In §2, we first show that if $\mathbf{T}(p,q)$ has a $\frac{1}{2}\mathbb{Z}^3$-primitive subdivision, then, after a preliminary normalization, this subdivision must contain two special edge paths, $X_1X_2\ldots X_{q-1}$ and $X_{\pi(1)}X_{\pi(2)}\ldots X_{\pi(q-1)}$, which we call "maximal chains". Here, the X_i's denote, in a certain specific order, the $q-1$ points of the lattice $\frac{1}{2}\mathbb{Z}^3$ which lie in the interior of the "central parallelogram" (see Figure 2) of $\mathbf{T}(p,q)$, and π is a certain permutation of $\{1,2,\ldots,q-1\}$ depending only on p and q. So it is necessary that the intersection of the two maximal chains be a common subcomplex. The "only if" part of Theorem 1 follows by checking that this obstruction is nontrivial unless $p=1$ or $q-1$. In §3 we complete the proof of Theorem 1 by constructing, using maximal chains, a $\frac{1}{2}\mathbb{Z}^3$-primitive subdivision of $\mathbf{T}(1,q) \simeq \mathbf{T}(q-1,q)$.

It turns out that the same construction gives a $\frac{1}{2}\mathbb{Z}^3$-elementary subdivision of any $\mathbf{T}(p,q)$, if we use instead any two "chains", i.e., edge paths determined by vertex subsequences $X_{i_1}X_{i_2}\ldots X_{i_t}$ and $X_{\pi(j_1)}X_{\pi(j_2)}\ldots X_{\pi(j_s)}$, but again such that their intersection is a common subcomplex. This gives a simple proof of the result of Knudsen, Mumford and Waterman, after which we insert an example of a non-convex polytope with six integral vertices which does not admit a $\mathbb{Z}^3$-elementary simplicial subdivision, to show exactly where convexity was used. If an "almost maximal" (see §3 for the definition) pair of chains is used, each non-primitive $\frac{1}{2}\mathbb{Z}^3$-elementary tetrahedron of this subdivision is isomorphic to a $\frac{1}{2}\mathbf{T}(1,t)$, and using Theorem 1 we can subdivide again to obtain a $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision. So, the rest of §3 is devoted to constructing various pairs of almost maximal chains whose intersection is a common subcomplex. First, the special but quite interesting constructions of Proposition 7, and the ensuing Remark 7, which already give numerous instances of $\frac{1}{4}\mathbb{Z}^3$-primitive subdivisions, and then the general "quadrant construction" which yields Theorem 2.

Our problem belongs to 3-dimensional affine geometry, and fittingly, the solution we have given below is also pleasingly geometrical; *except* for the fact that we make an essential use of the theorem of White, and all proofs of this purely geometrical result seem to involve some number theory; also, no satisfactory higher dimensional analogue of this result is known. Before starting work we recall below some standard simplicial definitions (for more see e.g., [**Hu**]).

Notations. The convex hull of a finite set of points is called a *convex cell,* the extreme points being its *vertices*. If these happen to be affinely independent the cell becomes a *simplex*, i.e., since we work in 3-space, a point, edge, triangle or tetrahedon. A finite set K of simplices (likewise of cells) with pairwise disjoint interiors, constitutes a *complex* if the faces of each $\sigma \in K$ are also in K. Then we say that K is a linear *triangulation* of the space $|K|$ obtained by taking the union of these σ's. A simplicial complex L with $|L| = |K|$ is a linear simplicial *subdivision* of K if each simplex of L is contained in a simplex of K. An *edge path* is a one-dimensional simplicial complex, with edges oriented and totally ordered in such a way that the initial vertex of each coincides with the final vertex of the preceding. The *open star* $\mathrm{St}_K(\sigma)$ of a simplex $\sigma \in K$ consists of all open simplices of K having σ as a face, and the *link* $\mathrm{Lk}_K(\sigma)$ of σ consists of all simplices disjoint from σ which join with σ to form simplices of K. We sometimes identify a simplex with the complex obtained by adjoining all its faces, and a simplicial complex with the space it triangulates.

2. Obstruction.

In this section we analyze linear $\frac{1}{2}\mathbb{Z}^3$-primitive subdivisions of $\mathbf{T}(p, q)$ and thereby prove the "only if" part of Theorem 1. Since this is obviously true for $q \leq 4$, we can assume that $q \geq 5$. The strategy roughly is to reduce the 3-dimensional problem concerning the join $AB * CD$ to a 2-dimensional problem concerning the cartesion product $AB \times CD$.

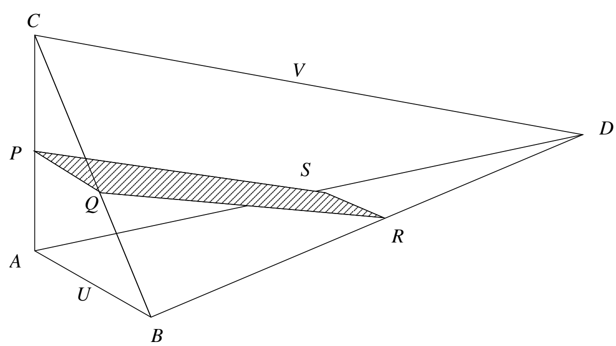


Figure 2.

CENTRAL PARALLELOGRAM. In $\mathbf{T}(p, q)$, the points of the lattice $\frac{1}{2}\mathbb{Z}^3$, other than the four vertices A, B, C, D, and the six midpoints P, Q, R, S, U, V — see Figure 2 — of the edges AC, BC, BD, AD, AB, CD respectively, must all be located in the interior of the parallelogram $PQRS$, i.e., the section of the tetrahedron $\mathbf{T}(p,q)$ by the plane $z = 1/2$. Note that this parallelogram has area $q/4$.

THE LATTICE POINTS ON THE PARALLELOGRAM. If we draw lines (Figure 3 shows the case $p = 3,\ q = 5$) parallel to PQ in the plane $z = \frac{1}{2}$, at equal distances 1/2 from each other, then (lengths equal to 1/2 of) $q-1$ of these lines pass through the interior of our parallelogram, so it follows that *there are* $q-1$ *points of* $\frac{1}{2}\mathbb{Z}^3$ *in the interior of the parallelogram,* one on each of these $q-1$ lines (which will sometimes be called HORIZONTALS). We denote these points by $X_1, \ldots, X_{q-1}$ in increasing order of their distance from PQ. So, X_1 is the point nearest to PQ, while X_{q-1} is the farthest, i.e., the one nearest to RS.

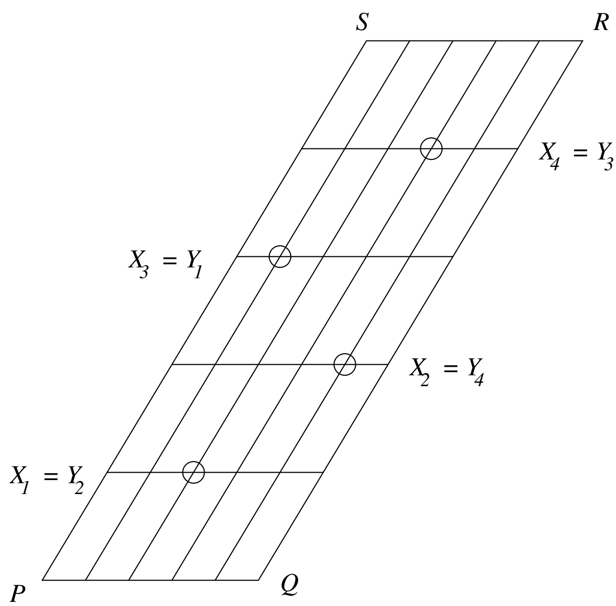


Figure 3.

Likewise, we can subdivide $PQRS$ into q parallelograms, each of area 1/4, by drawing $q-1$ lines parallel to SP, and at equal distances from each other, and precisely one of these points is found on each of these $q-1$ parallels (these will sometimes be called VERTICALS). We will sometimes denote the same set of $q-1$ points (of $\frac{1}{2}\mathbb{Z}^3$ in the interior of $PQRS$) by $Y_1, \ldots Y_{q-1}$ in increasing order of their distance from SP. So, now Y_1 is the point nearest to SP, while Y_{q-1} is the farthest, i.e., the one nearest to QR. We define the PERMUTATION $\pi = \pi(p,q)$ of $\{1, 2, \ldots, q-1\}$ by $Y_j = X_{\pi(j)}$, and, as mentioned already in §1, it is this permutation which will play the leading role in the following:

NORMALIZATION. We begin by showing that we may consider only subdivisions well-behaved with respect to the parallelogram.

Proposition 1. *If the elementary tetrahedron* $\mathbf{T} = \mathbf{T}(p, q)$, $q \geq 5$, *admits a linear simplicial subdivision* $\mathbf{K} = \mathbf{K}(p, q)$ *which is primitive with respect to* $\frac{1}{2}\mathbb{Z}^3$, *then it also admits another in which the central section occurs as the space* $|\mathbf{P}(p, q)|$ *of a subcomplex* $\mathbf{P}(p, q) \subset \mathbf{K}(p, q)$.

Proof. A simplicial subdivision of our tetrahedron fails to have the parallelogram as (the space of) a subcomplex iff it contains an edge having one vertex below it and the other above it. On the boundary of the tetrahedron there are only four possibilities for such an edge, namely UC, UD, VA, VB (e.g., AC is not a possibility because it has P on its interior). In case some or all of these four edges occur in $\mathbf{K}$ we will first get rid of them as follows:

Suppose, e.g., that UC is an edge of $\mathbf{K}$. Then its two incident boundary triangles, which have to be UPC and UQC, must be triangles of $\mathbf{K}$. The two elementary tetrahedra $UPCX_1$ and $UQCX_1$ both have volume 1/48, so are primitive with respect to $\frac{1}{2}\mathbb{Z}^3$, and all other possible tetrahedra $UPCZ$, $UQCZ$, $Z = V, S, R, X_i$, $i \geq 2$, have bigger volumes. So these two tetrahedra must also be in $\mathbf{K}$. We now modify $\mathbf{K}$ by replacing these two tetrahedra, and their common face UCX_1, by the two primitive tetrahedra $CPQX_1$ and $UPQX_1$, and their common face PQX_1. This retriangulation of the *open star* in $\mathbf{K}$ of the edge UC leaves the triangulations on the other 3 faces of our tetrahedron unaltered. Likewise, if say, $BV \in \mathbf{K}$, then $BVQY_{q-1} \in \mathbf{K}$ and $BVRY_{q-1} \in \mathbf{K}$, and we can get rid of BV by replacing these two tetrahedra and their common face by the primitive tetrahedra $QRY_{q-1}B$, $QRY_{q-1}V$ and their common face, etc.

In case q (and so also $p+1$) *is even*, our modified $\mathbf{K}$, which now contains none of the four edges UC, UD, VA, VB, has no edge joining a vertex below the parallelogram to one above, and so must contain a subcomplex $\mathbf{P}$ of the required kind. To see this note that UV is not a possible edge of $\mathbf{K}$, because its mid-point $M = \frac{1}{4}(p+1, q, 2) \in \frac{1}{4}\mathbb{Z}^3$ — i.e., the BARYCENTER M of $\mathbf{T}(p, q)$, i.e., the intersection of the diagonals of the central section $PQRS$ — belongs to the lattice $\frac{1}{2}\mathbb{Z}^3$ in this case.

In case q *is odd*, the barycenter M is not contained in $\frac{1}{2}\mathbb{Z}^3$, and UV remains a possible edge of $\mathbf{K}$ joining a vertex below the parallelogram to one above. We now describe a procedure for getting rid of UV.

First, note that if $UV \in \mathbf{K}$, and $UVIJ$ is any tetrahedron of $\mathbf{K}$ incident to this edge, then I and J must be both on the parallelogram. Indeed, neither I nor J can be C, D, A or B, for otherwise, one of the 4 already excluded edges UC, UC, VA, VB would be in $\mathbf{K}$. The union $\partial\mathbf{M}$ of all such edges IJ, i.e., the *link* of UV in $\mathbf{K}$, bounds a polygonal region $\mathbf{M}$ — namely the union of all the triangles MIJ — of the parallelogram, which meets $\frac{1}{2}\mathbb{Z}^3$ only in its vertices. We equip $\mathbf{M}$ with any triangulation which is elementary (so with triangles of area 1/8) with respect to the half integral lattice of the plane $z = 1/2$ (see Lemma 1 below). We then modify $\mathbf{K}$ by replacing the tetrahedra $UVIJ$, and their common faces, by all the simplices obtained by coning the triangulated $\mathbf{M}$ over U and V.

This primitive retriangulation of the open star of UV serves to get rid of the edge UV. Since the new $\mathbf{K}$ has no edge joining a vertex below the parallelogram to one above, it will have a subcomplex $\mathbf{P}$ which covers this parallelogram. □

Remark 1. *We can further modify* $\mathbf{K}(p, q)$ *so that its restriction to each bounding face is* STANDARD (see Figure 4): I.e., that it is cut up into four smaller triangles by joining the mid-points of its sides to each other. (While subdividing simplicial complexes, as in §3, this finesse is useful, because the subdivisions of two tetrahedra sharing a common face will now match on this common face.) For example, if PBX_1Q and PBX_1U are in $\mathbf{K}(p, q)$, we retriangulate the open star of their common face PBX_1 by using QUX_1P and QUX_1B and their common face QUX_1.

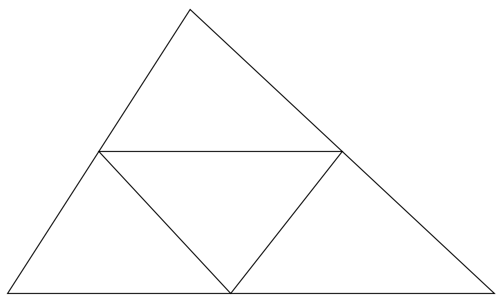


Figure 4.

We insert here a proof of the following which we use in quite a few arguments, e.g., the one just made above. This is an essentially 2-dimensional result, its 3-dimensional analogue is false (see §3).

Lemma 1. *Any* (possibly non-convex) $\mathbb{Z}^2$*-elementary polygonal region* $\mathbf{M}$ *of* $\mathbb{R}^2$ *can be subdivided into* $\mathbb{Z}^2$*-elementary triangles.*

Proof. We use induction on the number n of the vertices $\ldots v_{i-1}v_iv_{i+1}\ldots$ of the bounding polygon. If $n = 3$ there is nothing to do. If $n \geq 4$, the sum of the n polygonal angles being $(n-2)\pi$, we can choose a vertex v_i such that the angle $v_{i-1}v_iv_{i+1}$ is less than π. If the triangle $v_{i-1}v_iv_{i+1}$ is elementary, we just add it to any elementary subdivision of the region bounded by the polygon $\ldots v_{i-2}v_{i-1}v_{i+1}\ldots$ having $n-1$ vertices. Otherwise, we take a $u \in \mathbb{Z}^2$, $u \neq v_{i-1}, v_i, v_{i+1}$ in this triangle, which is nearest to v_i. Then u must be a vertex of $\mathbf{M}$, with the interior of the edge uv_i contained in the interior of $\mathbf{M}$. Now, using induction, we subdivide the 2 regions into which $\mathbf{M}$ is separated by this edge. □

BOTTOM HALF. We plan next to use the following lemma to derive some conditions on the subcomplex $\mathbf{P}(p,q)$ of Proposition 1 which follow *solely* by using the primitivity, with respect to $\frac{1}{2}\mathbb{Z}^3$, of the tetrahedra of $\mathbf{K}(p,q)$ which cover the portion $z \le 1/2$ of our tetrahedron. This triangulation of the bottom half will be denoted $\mathbf{K}_X(p,q)$.

Lemma 2. *If HIJ is a $\frac{1}{2}\mathbb{Z}^3$-elementary triangle on the central parallelogram the* 3 *$\frac{1}{2}\mathbb{Z}^3$-elementary tetrahedra $AHIJ$, $UHIJ$, $BHIJ$ are all $\frac{1}{2}\mathbb{Z}^3$-primitive. Furthermore, if IJ is an $\frac{1}{2}\mathbb{Z}^3$-elementary edge on the central parallelogram, with $I = \frac{1}{2}(i_1, i_2, 1)$, $J = \frac{1}{2}(i_1, i_2, 1)$, $j_2 \ge i_2$, then the $\frac{1}{2}\mathbb{Z}^3$-elementary tetrahedra $AUIJ$ and $UBIJ$ are isomorphic to each other, and have the standard model $\frac{1}{2}\mathbf{T}(j_1 - i_1, j_2 - i_2)$, $j_2 \ge i_2$. So these tetrahedra are $\frac{1}{2}\mathbb{Z}^3$-primitive if and only if IJ is one of the $q+2$ edges:*

$$(*) \qquad PX_1, QX_1; X_1X_2, X_2X_3, \ldots, X_{q-2}X_{q-1}; X_{q-1}S, X_{q-1}R.$$

Proof. Because the area of HIJ is $1/8$, the volume of $AHIJ$, $UHIJ$ or $BHIJ$ is $1/3 \times 1/8 \times 1/2 = 1/48$, which shows the first part.

Next note (cf. Figure 3) that the first coordinate of X_{k+1} is either the same as that of X_k, or $\frac{1}{2}$ more. Therefore, $j_2 \ge i_2$ always implies $0 \le j_1 - i_1 \le j_2 - i_2$. Further, IJ has no point of $\frac{1}{2}\mathbb{Z}^3$ other than its vertices, iff $j_1 - i_1$ and $j_2 - i_2$ are relatively prime to each other.

For the second part we first give an affine linear transformation which preserves $\frac{1}{2}\mathbb{Z}^3$ and maps $AUIJ$ onto $UBIJ$, viz., the SHEAR which keeps all points of the plane $z = \frac{1}{2}$ fixed, and maps each parallel plane $z = \frac{1}{2}(1-k)$ onto itself by adding $\frac{1}{2}k$ to the first coordinate of its points. Likewise, the $\frac{1}{2}\mathbb{Z}^3$-preserving shearing transformation which keeps the plane $z = 0$ fixed, and translates I to $P = \frac{1}{2}(0,0,1)$, moves the point J to $\frac{1}{2}(j_1 - i_1, j_2 - i_2, 1)$, and thus images the tetrahedron $AUIJ$ onto the standard model $\frac{1}{2}\mathbf{T}(j_1 - i_1, j_2 - i_2)$.

It follows that the necessary and sufficient condition for primitivity of these tetrahedra is $j_2 - i_2 = 1$, which happens if and only if IJ is one of the edges listed as $(*)$. □

Remark 2. Alternatively, the volume $\frac{1}{6}(AU \times AJ).AI$ of $AUIJ$, and the volume $\frac{1}{6}(UB \times UJ).UI$ of $UBIJ$, are both equal to $1/48(j_2 - i_2)$, being the value of the equal determinants

$$\frac{1}{48}\begin{vmatrix} 1 & 0 & 0 \\ j_1 & j_2 & 1 \\ i_1 & i_2 & 1 \end{vmatrix} = \frac{1}{48}\begin{vmatrix} 1 & 0 & \\ j_1 - 1 & j_2 - 1 & 1 \\ i_1 - 1 & i_2 & 1 \end{vmatrix}.$$

So the necessary and sufficient condition for primitivity is $j_2 - i_2 = 1$, which happens iff IJ is one of the edges $(*)$. This alone will be used in the present section; however in §3 we use the fact that these elementary tetrahedra have "p"$= j_1 - i_1$ and "q"$= j_2 - i_2$.

MAXIMAL CHAIN. The edge path $\mathcal{X} = X_1X_2\dots X_{q-2}X_{q-1}$ is called the maximal X-chain (likewise $\mathcal{Y} = Y_1Y_2\dots Y_{q-2}Y_{q-1}$ will be called the maximal Y-chain); note that $(*)$ of Lemma 2 consists of its $q-2$ X-EDGES X_iX_{i+1}, together with four more: Two each stuck to each end X_1 and X_{q-1}. For small values of (p,q) one can quickly plot the lattice points of the parallelogram on quadrille paper and then draw $\mathcal{X}$: Figure 5 shows this for the two cases $(p,q) = (4,7)$ and $(3,8)$. We will refer to PQX_1 and $X_{q-1}RS$ as X-END TRIANGLES, and call the remaining two regions of the parallelogram LEFT X-POLYGONAL and RIGHT X-POLYGONAL according as it is to the left or right of the chain as we move on it from X_1 to X_{q-1}. Other usages of "left" and "right" will be compatible to the one just made; also, the suffix "X-", which is used to distinguish these concepts from their obvious "Y-" analogues will often be omitted.

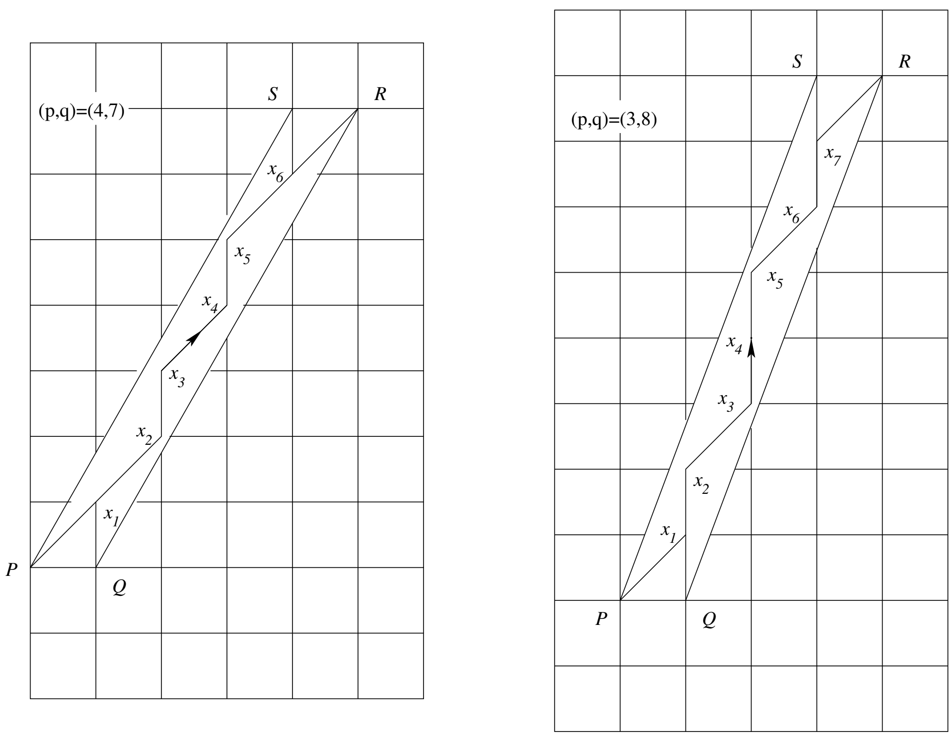


Figure 5.

WEIGHT. An internal edge IJ of the subcomplex $\mathbf{P}(p,q)$ of $\mathbf{K}_X(p,q)$, a $\frac{1}{2}\mathbb{Z}^3$-primitive simplicial subdivision of the bottom half, will be said to have X-weight 0, 1 or 2, depending on whether none, one, or both of the tetrahedra $AUIJ$ and $UBIJ$ are in $\mathbf{K}_X(p,q)$. The following three facts about weight will be important:

(Wt0) The weight of IJ is 0 iff its two incident triangles in $\mathbf{P}(p,q)$ cone over the same vertex from $\{A, U, B\}$ to give tetrahedra of $\mathbf{K}_X(p,q)$. For "only if" assume, e.g., that $IJU \in \mathbf{K}_X(p,q)$. Since the weight of IJ is 0, the fourth vertices, of the two tetrahedra of $\mathbf{K}_X(p,q)$ incident to IJU, must be on the parallelogram, i.e., must be the vertices H_l and H_r - respectively to the *left* and *right* of IJ — such that IJH_l, $IJH_r \in \mathbf{P}(p,q)$. The converse "if" is obvious.

(Wt1) The weight of IJ is 1, with say $IJAU$ in $\mathbf{K}_X(p,q)$ (the statement for $IJUB \in \mathbf{K}_X(p,q)$ is analogous), *iff* (with the same notation as before) *IJH_lA and IJH_rU are in* $\mathbf{K}_X(p,q)$. For "only if" note that the open tetrahedron $IJAU$ intersects both IJH_rA and IJH_lU. So the fourth vertex of the other tetrahedron incident to IJA has to be H_l. Again, since weight of IJ is 1, the fourth vertex of the other tetrahedron incident to IJU is also on $\mathbf{P}(p,q)$, and so must be H_r. For the converse "if" note that, for the fourth vertex of the other tetrahedron incident to IJA, there is only one choice, namely U.

(Wt2) The weight of IJ is 2 iff AJH_lA and AJH_rB are in $\mathbf{K}_X(p,q)$. The verification is similar, so can be omitted.

Proposition 2. *For any $\frac{1}{2}\mathbb{Z}^3$-primitive simplicial subdivision $\mathbf{K}_X(p,q)$ of the bottom half, $z \leq 1/2$, of $\mathbf{T}(p,q)$, $q \geq 5$, the subcomplex $\mathbf{P}(p,q)$ covering the parallelogram must contain all the $q+2$ edges $(*)$, and each of the $q-2$ edges X_iX_{i+1} must have weight 2 in $\mathbf{P}(p,q)$.*

Proof. Since the area of the triangle PQZ is bigger than $1/8$ if $Z = S$, R, or X_i, the edge $PQ \in \mathbf{P}(p,q)$ is incident to only one elementary triangle of the parallelogram, namely the end triangle PQX_1, and so we must have $PQX_1 \in \mathbf{P}(p,q)$. Likewise, $X_{q-1}RS \in \mathbf{P}(p,q)$.

The total weight of the edges must be $2q$. To see this note, since A, U, B are collinear, that any tetrahedron of $\mathbf{K}_X(p,q)$ is either of the FIRST KIND, with 3 vertices on the parallelogram, or of the SECOND KIND, with just 2 on the parallelogram, and the total weight is the same as the number of tetrahedra of the second kind. The total volume of all tetrahedra of the first kind is the same as the volume of the pyramid of $PQRS$ over B, i.e., $q/24$. (Thus the number of tetrahedra of the first kind, i.e., the number of triangles of $\mathbf{P}(p,q)$, is also $q/24 \div 1/48 = 2q$.) The remaining volume, i.e., the volume of all tetrahedra of the second kind, is the same as that of the tetrahedron $PABS$, which is also $q/24$. This shows that the total number of tetrahedra of the second kind is $q/24 \div 1/48 = 2q$.

By Lemma 2 we know all edges which could occur with weight ≥ 1. So we know (see Figure 5) that, if one of the edges X_iX_{i+1} is not in $\mathbf{K}_X(p,q)$, or else, even if all these edges are in $\mathbf{K}_X(p,q)$, but one of them, X_iX_{i+1}, has weight 0, then we can "join" any 2 triangles of $\mathbf{P}(p,q)$, other than the 2 end triangles, by means of a sequence of triangles, such that each shares a weight zero edge with the previous one. Using (Wt0), these triangles must all join to the same vertex from $\{A, U, B\}$ and so there are at most 4 edges (of the end triangles) of positive weight. Since $q \geq 5$ the total weight would be less than $2q$. The entire X-chain must thus be in the subcomplex $\mathbf{P}(p, q)$ and all its edges must have positive weight in $\mathbf{K}_X(p, q)$.

In case one of its edges X_iX_{i+1} is only of weight 1, say with $AUX_iX_{i+1} \in \mathbf{K}_X(p, q)$, then by (Wt1), the left triangle of $\mathbf{P}(p, q)$ incident to X_iX_{i+1} joins with A, and the right triangle with U, to form tetrahedra of $\mathbf{K}_X(p, q)$. Using again the argument of the previous paragraph, it follows that all triangles of the left polygonal region must join with A to give tetrahedra of $\mathbf{K}_X(p, q)$, and likewise, all triangles of the right polygonal region must join with U to form tetrahedra of $\mathbf{K}_X(p, q)$. Using the converse part of (Wt1) all edges of the chain must have weight 1. Since the 2 edges stuck at each end can now contribute at most weight 3, we see that total weight is at most $(q - 2) + 2.3 = q + 4$ which is less than the required $2q$ because $q \geq 5$. It follows that all edges X_iX_{i+1} have weight 2. $\square$

Remark 3. In $\mathbf{K}_X(p, q)$ the end triangle PQX_1, respectively $X_{q-1}RS$, is incident to one of the tetrahedra from $\{APQX_1, UPQX_1, BPQX_1\}$, respectively $\{AX_{q-1}RS, UX_{q-1}RS, BX_{q-1}RS\}$. In case the boundary faces are subdivided in the standard way (see Remark 1) these clearly must be given by the second alternative, i.e., $UPQX_1 \in \mathbf{K}_X(p, q)$ and $UX_{q-1}RS \in \mathbf{K}_X(p, q)$. Also, using (Wt2) and (Wt0), all triangles of the left polygonal region of $\mathbf{P}(p, q)$ must cone over A, and all those of the right polygonal region over B. Note that now each of the internal edges of an end triangle has weight 1 (if boundary subdivision of that face is not standard one of them will have weight 2 and the other weight 0). Thus, with boundary subdivision standard, *the triangulation* $\mathbf{K}_X(p, q)$ *is unique,* the possible ambiguity being only as to how one chooses to elementarily subdivide the left and right polygonal regions of the parallelogram. [In fact the parallelogram has a unique triangulation with X a subcomplex, thus there is no ambiguity.]

Remark 4. The above arguments made only a very mild use of the half-integrality of the vertices X_i. They show in fact that *all the conclusions of Proposition* 1 *hold even if the* $q - 1$ *vertices* X_i *are any interior points of* $PQRS$*, one on each horizontal, provided all the tetrahedra of the subdivision have volume* 1/48.

UPPER HALF. The next result gives the analogous conditions, imposed solely by the primitivity, with respect to $\frac{1}{2}\mathbb{Z}^3$, of the tetrahedra of $\mathbf{K}(p, q)$ which cover the portion $z \geq 1/2$ of our tetrahedron, and can be established by mimicking the proof of Proposition 2. However we give below another proof, in the course of which we will define and use an interesting SYMMETRY ϕ between the two halves.

Proposition 3. *For any $\frac{1}{2}\mathbb{Z}^3$-primitive simplicial subdivision $\mathbf{K}_Y(p,q)$ of the upper half, $z \geq 1/2$, of $\mathbf{T}(p,q)$, $q \geq 5$, the subcomplex $\mathbf{P}(p,q)$ covering the parallelogram must contain all the $q+2$ edges:*

$$(**) \qquad PY_1, SY_1; Y_1Y_2, Y_2Y_3, \ldots, Y_{q-2}Y_{q-1}; Y_{q-1}Q, Y_{q-1}R.$$

Furthermore, each of the $q-2$ Y-edges $Y_iY_{i+1} \in \mathcal{Y}$ has Y-weight 2, i.e., CVY_iY_{i+1} and VDY_iY_{i+1} must both be in $\mathbf{K}_Y(p,q)$.

For example, for the same values of (p,q) as were used to draw Figure 5, the maximal Y-chains $\mathcal{Y}$ are as shown in Figure 6 below (these chains can be made to look less compressed by, say doubling the scale in the x-direction).

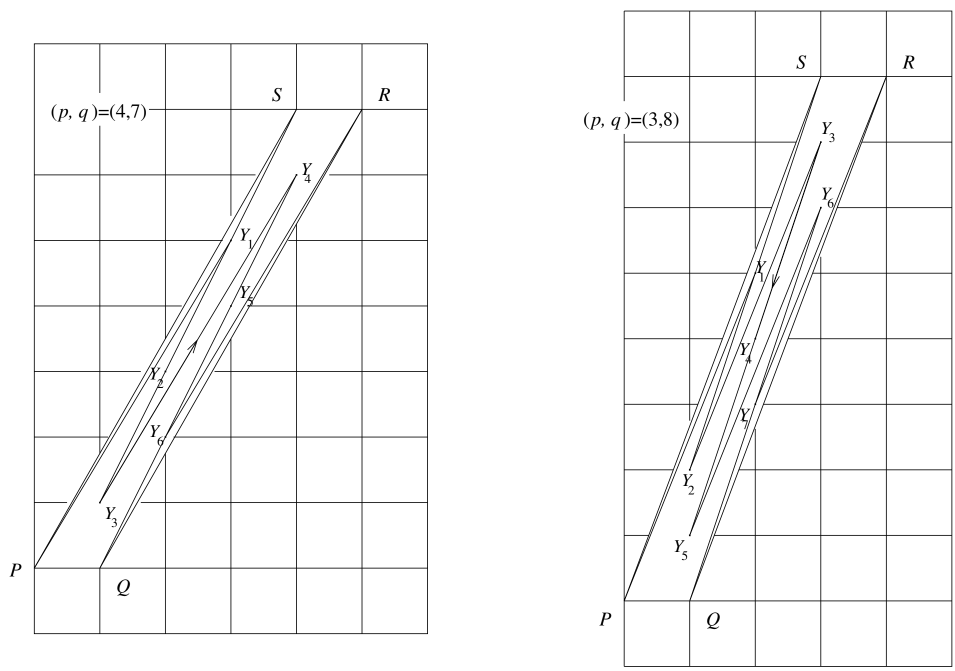


Figure 6.

Proof. The aforementioned ϕ will be the linear transformation of 3-space, having the barycenter M as its fixed point, and mapping A, B and C respectively to C, D and B. Since MA, MB, MC and MD add up to the zero vector, it follows that ϕ also maps D to A. Since ϕ preserves the tetrahedron, it must be volume preserving. Moreover, it switches the upper and bottom halves, interchanging the point V with U, and "rotates" the parallelogram, mapping P, Q, R and S respectively to Q, R, S and P. Its restriction to this parallelogram being area preserving, it follows that ϕ must map the straight line through Y_i parallel to SP onto the straight line through X_i parallel to PQ (however we note that ϕ need not map the point Y_i to the point X_i).

Applying ϕ to the given $\mathbf{K}_Y(p,q)$ we obtain a linear subdivision of the bottom half, having P, Q, R, S, A, U, B and $\phi(Y_i)$, $1 \leq i \leq q-1$, as its vertices, with areas of all triangles on the parallelogram 1/8, and volumes of all tetrahedra 1/48. The result now follows by Remark 4 which tells us that all the conclusions of Proposition 1 apply to this combinatorially isomorphic subdivision of the bottom half. □

THE OBSTRUCTION. As Figures 5 and 6 suggest, the set theoretic intersection $\mathcal{X} \cap \mathcal{Y}$ is seldom a subcomplex, and so, it is rarely the case that *both* $\mathcal{X}$ and $\mathcal{Y}$ are subcomplexes of $\mathbf{P}(p,q)$. More precisely, the following concludes the proof of the "only if" part of Theorem 1:

Proposition 4. *An X-edge of the parallelogram is a Y-edge iff $p = 1$ or $q-1$ and then all X-edges are Y-edges. In all other cases there is a pair (X-edge, Y-edge) with intersection a non half-integral point.*

Proof. All points being now on the parallelogram, their omitted third coordinate will always be 1/2. The segment with end points $I = \frac{1}{2}(i_1, i_2)$, $J = \frac{1}{2}(j_1, j_2) \in \frac{1}{2}\mathbb{Z}^3$ is an X-edge iff their second coordinates differ by one, say $j_2 - i_2 = 1$. Note that the line through X_i parallel to SP has equation $qx - py = k$ where $1 \leq k \leq q-1$ is the integer such that $X_i = Y_k$. So IJ is also a Y-edge iff $(qj_1 - pj_2) - (qi_1 - pi_2) = 1$ or -1. Because $1 \leq p < q$, the first case, $q(j_1 - j_2) = p + 1$, occurs iff $q = p+1$ and $j_1 - i_1 = 1$, and the second case, $q(j_1 - i_1) = p - 1$, happens iff $p = 1$ and $j_1 - i_1 = 0$. *In case $p = q-1$ one has $X_t = \frac{1}{2}(t,t) = Y_t$, $1 \leq t < q-1$, and in case $p = 1$ one has $X_t = \frac{1}{2}(1,t) = Y_{q-t}$: So in either of these cases the X-chain and Y-chain coincide and are straight lines.* As we will see, in all other cases, both the chains are crooked.

Case $2p < q$. The slope of the line SP being q/p — see Figure 7 — we will have $X_1 = \frac{1}{2}(1,1), X_2 = \frac{1}{2}(1,2), \ldots, X_t = \frac{1}{2}(1,t)$, where t denotes the largest integer such that $tp < q$, and then, since $(t+1)p > q$ (as g.c.d. of $p \neq 1$ and q is 1 we cannot have $(t+1)p = q$) the next $X_{t+1} = \frac{1}{2}(2, t+1)$. So the X-chain is crooked with "first bend" at X_t. *Just after this "first bend" of the X-chain, there is a non half-integral intersection with a Y-edge.* To see this note that $X_1 = (1,1) = Y_{q-p}$ is neither Y_1 nor Y_{q-1}, so the Y-chain too is crooked with a bend at this vertex. Also $X_{t+1} = \frac{1}{2}(2, t+1) = Y_{q2-p(t+1)}$ comes after this vertex in the Y-chain because $q2 - p(t+1) > q - p$ is the same as $q > pt$, and $X_2 = \frac{1}{2}(1,2) = Y_{q-p2}$ comes before Y_{q-p-1} because $q - p2 < q - p - 1$ is the same as $p > 1$. So Y_{q-p-1} is trapped in the indicated shaded region and $Y_{q-p-1}Y_{q-p}$ must intersect X_tX_{t+1} at an interior point.

Case $2p > q$.[3] The slope of QR being q/p we now have $X_1 = \frac{1}{2}(1,1), X_2 = \frac{1}{2}(2,2), \ldots, X_t = \frac{1}{2}(t,t)$, where t is the largest integer such that $t/(t-1) > q/p$. We can have $(t+1)/t = q/p$, i.e., $p(t+1) = qt$, only if $p = t = q-1$,

[3] This in fact follows from the case $2p < q$ because of Lemma 3 below.

which we ruled out; so $(t+1)/t > q/p$ and the next $X_{t+1} = \frac{1}{2}(t, t+1)$. Again, there is an intersection with a Y-edge just after this "first bend" of the X-chain. This time it is Y_{q-p+1} which is trapped in the shaded region, and so $Y_{q-p}Y_{q-p+1}$ intersects X_tX_{t+1}. This because $X_{t+1} = \frac{1}{2}(t, t+1) = Y_{qt-p(t+1)}$ comes before $X_1 = Y_{q-p}$ on the Y-chain as $qt - p(t+1) < q - p$ is the same as $q(t-1) < pt$, and Y_{q-p+1} is before X_2 as $q - p + 1 < 2q - p2$ is the same as $p < q - 1$. $\square$

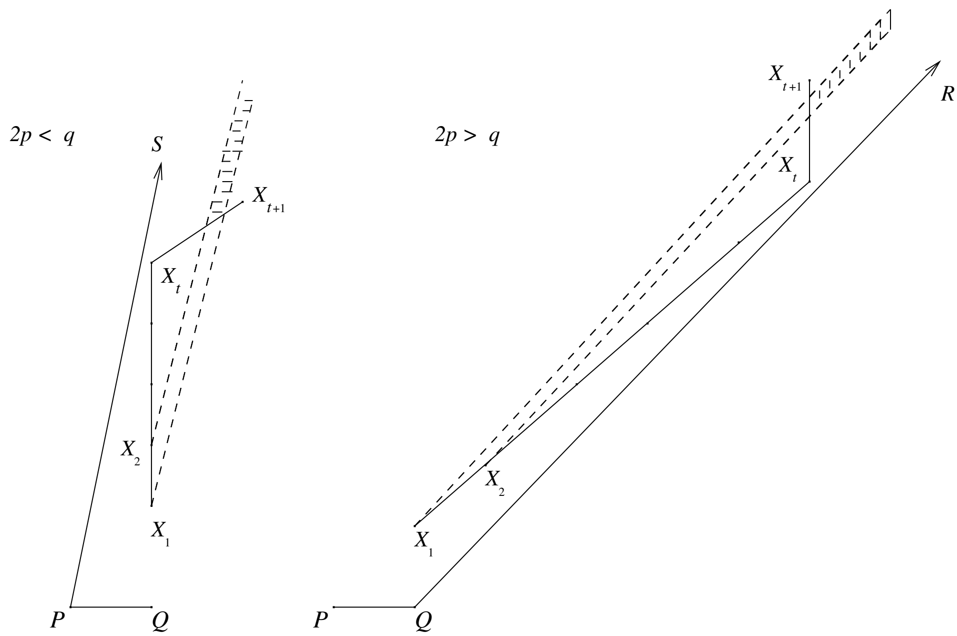


Figure 7.

Remark 5. In the course of the above proof we obtained a more geometric way of stating Theorem 1: *An $\mathbb{L}$-elementary tetrahedron admits a $\frac{1}{2}\mathbb{L}$-primitive subdivision iff the points of $\frac{1}{2}\mathbb{L}$ lying in its interior are collinear.* Note also that, even though the order 4 volume preserving affine transformation ϕ is not lattice preserving, the involution ϕ^2 is (and is in general — see proof of Lemma 3, §3 — the sole nonidentity automorphism of $\mathbf{T}(p, q)$). It preserves both halves, and, restricted to the central section, reflects it through M. *The maximal X- and Y-chains are thus preserved by reflection through the barycenter M*, which thus must be on both chains. If q is odd, $M \notin \frac{1}{2}\mathbb{Z}^3$, which gives, for this case, another proof of the second part of Proposition 4.

3. Construction.

The analysis of §2 already indicates — see Remark 3 — what a $\frac{1}{2}\mathbb{Z}^3$-primitive subdivision of either half should look like. We now construct such subdivisions. It turns out that the same construction also gives analogous $\frac{1}{2}\mathbb{Z}^3$-elementary, but not $\frac{1}{2}\mathbb{Z}^3$-primitive subdivisions if we employ non-maximal chains. These are defined as follows:

CHAINS. By an X-chain $\mathcal{C}$ we will mean an edge path $X_{i_1}X_{i_2}\dots X_{i_r}$, $1 = i_1 < i_2 < \cdots < i_r = q-1$, such that each of its edges, $X_{i_t}X_{i_{t+1}}$, has no half-integral points in its interior. Its vertices will usually be denoted $\overline{X}_t = X_{i_t}$, note that the first vertex $\overline{X}_1$ is always X_1, and the last $\overline{X}_r$ always X_{q-1}. The end triangles and the left/right polygonal regions of $\mathcal{C}$ are defined as for the case of the maximal X-chain, and Y-chains $\mathcal{D}$ are defined analogously.

SLICING. Using segments $\overline{P}_t\overline{Q}_t$, parallel to PQ, through the vertices $\overline{X}_t$ of the given chain $\mathcal{C}$, the bottom half becomes the union of $r-1$ *inner slices*, i.e., those between the planar sections $\overline{P}_t\overline{Q}_tAB$ and $\overline{P}_{t+1}\overline{Q}_{t+1}AB$, $1 \le t < r$ and two *end slices*, between $PQAB$ and $\overline{P}_1\overline{Q}_1AB$, and between $\overline{P}_r\overline{Q}_rAB$ and $SRAB$. As Figure 8 shows, an inner slice subdivides as the union of the tetrahedron $\overline{X}_t\overline{X}_{t+1} * AB$ (which subdivides further into two by means of the triangle $\overline{X}_t\overline{X}_{t+1}U$) and the pyramids of two quadrilaterals over A and B. From the first end slice we will carve out the tetrahedron $PQ\overline{X}_1U$, leaving us with the pyramid of the quadrilateral $AU\overline{X}_1\overline{P}_1$ over P, and of $UB\overline{Q}_1\overline{X}_1$ over Q, which we will further cut into two parts each by means of the triangles $AP\overline{X}_1$ and $BQ\overline{X}_1$, respectively. The second end slice is subdivided similarly.

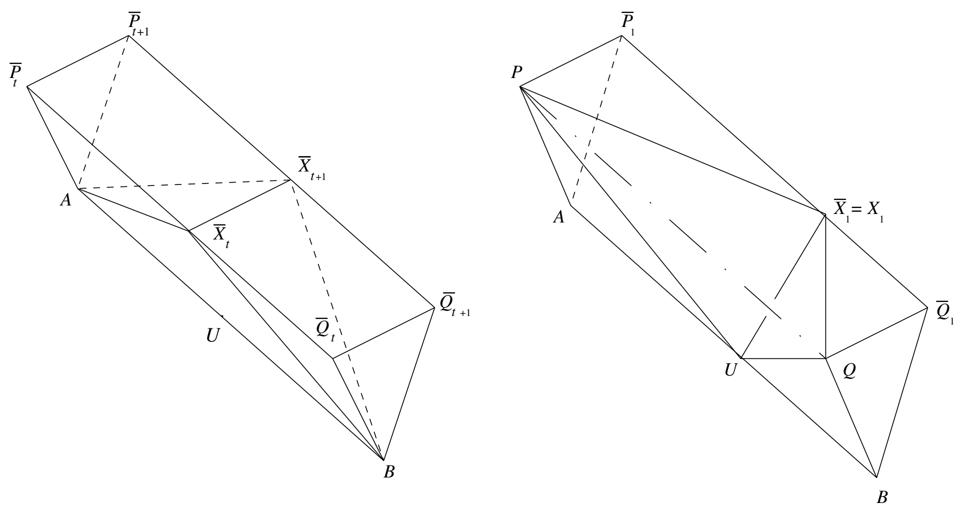


Figure 8.

Proposition 5. *For any X-chain $\mathcal{C}$ there is a linear simplicial $\frac{1}{2}\mathbb{Z}^3$-elementary subdivision $\mathbf{K}_{\mathcal{C}}(p,q)$ of the bottom half with the following tetrahedra:*

- (i) *Cones over A and B, respectively, of triangles of some elementary subdivisions of the left and right polygonal regions of $\mathcal{C}$,*
- (ii) *joins of edges of $\mathcal{C}$ with AU and UB, and*
- (iii) *$PAUX_1$, $PQUX_1$, $QUBX_1$ and $X_{q-1}SAU$, $X_{q-1}SRU$, $X_{q-1}RUB$.*

Proof. Putting the slices of Figure 8 back together we see that the entire bottom half is the union of (ii)′ the join of $\mathcal{C}$ with AB, (i)′ cones over A and B respectively of the left and right polygonal regions of $\mathcal{C}$, and the six tetrahedra (iii). The triangles $\overline{X}_t\overline{X}_{t+1}U$ subdivide (ii)′ into the tetrahedra (ii). Using Lemma 1 we now elementarily subdivide the left and right polygonal regions, and by coning these over A and B, subdivide (i)′ into the tetrahedra (i). □

Conclusion of proof of Theorem 1. Using Lemma 2 we see that the above tetrahedra (i) and (iii) are primitive, but *the tetrahedra of the second kind* (ii) *are all primitive if and only if $\mathcal{C} = \mathcal{X}$, the maximal X-chain.* Likewise for the upper half by virtue of the symmetry ϕ, thus *one always has primitive $\frac{1}{2}\mathbb{Z}^3$-subdivisions $\mathbf{K}_{\mathcal{X}}(p,q)$ and $\mathbf{K}_{\mathcal{Y}}(p,q)$ for the two halves separately.* However, Proposition 4 tells us that, unless $p = 1$ or $q-1$, $\mathcal{X} \cap \mathcal{Y}$ is not a subcomplex, thus $\mathbf{K}_{\mathcal{X}}(p,q)$ and $\mathbf{K}_{\mathcal{Y}}(p,q)$ cannot have the same $\mathbf{P}(p,q)$, and so cannot be put together to obtain a $\frac{1}{2}\mathbb{Z}^3$-primitive subdivision of $\mathbf{T}(p,q)$. If $p = 1$ or $q-1$, $\mathcal{X} = \mathcal{Y}$ (except for a change of direction in the $p = 1$ case). The subdivisions of the two halves can therefore be chosen with the same $\mathbf{P}(p,q)$, and then fit together to show the remaining "if" part of Theorem 1. □

For the sake of completeness we note here that, for the trivial *primitive case* $q = 1$ (now there are no X_i, so no $\mathcal{C}$) we can subdivide the bottom into two tetrahedra, and a pyramid over the central section, as in Figure 9, and cutting the parallelogram into two using either PR or QS, obtain a $\frac{1}{2}\mathbb{Z}^3$-subdivision.

$(\mathcal{C},\mathcal{D})$-SUBDIVISIONS. Given a pair (X-chain $\mathcal{C}$, Y-chain $\mathcal{D}$), *with $\mathcal{C} \cap \mathcal{D}$ a common subcomplex of both* $\mathcal{C}$ and $\mathcal{D}$ (note $\mathcal{C} \cap \mathcal{D} \neq \emptyset$ always) we can, using Lemma 1, find elementary subdivisions $\mathbf{P}(p,q)$ of the parallelogram containing both $\mathcal{C}$ and $\mathcal{D}$ as subcomplexes. Then, using Proposition 5, we can construct an $\frac{1}{2}\mathbb{Z}^3$-elementary subdivision $\mathbf{K}_{\mathcal{C}}(p,q) \supset \mathbf{P}(p,q)$ of the bottom half, and analogously, $\mathbf{K}_{\mathcal{D}}(p,q) \supset \mathbf{P}(p,q)$ of the top half. The resultant $\frac{1}{2}\mathbb{Z}^3$-elementary subdivision $\mathbf{K}_{\mathcal{C}}(p,q) \cup \mathbf{K}_{\mathcal{D}}(p,q)$ of $\mathbf{T}(p,q)$ will be called a $(\mathcal{C},\mathcal{D})$-subdivision and denoted by $\mathbf{K}_{\mathcal{C},\mathcal{D}}(p,q)$. Such pairs $(\mathcal{C},\mathcal{D})$ always exist, as we show in the course of the following simple proof of the (3-dimensional non-projective) Knudsen-Mumford-Waterman theorem.

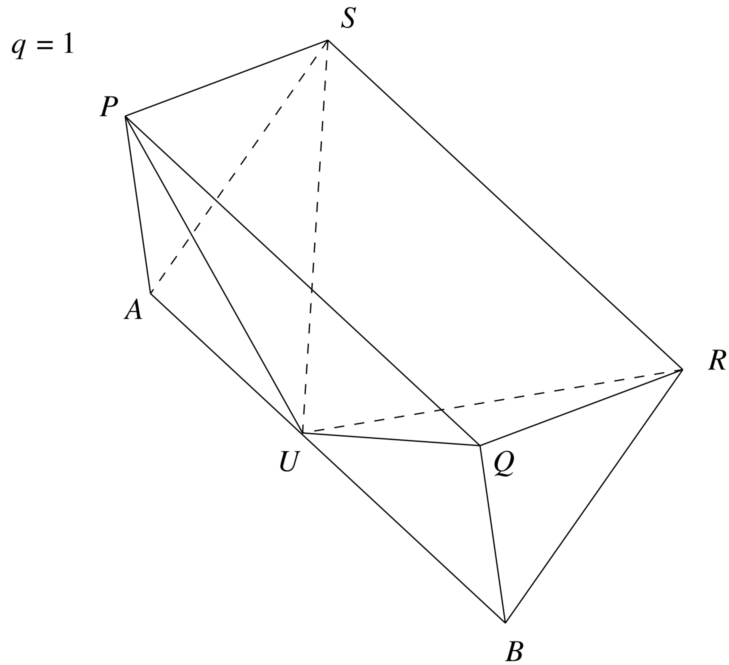


Figure 9.

Proposition 6. *Given any finite convex cell complex in* $\mathbb{R}^3$ *with all vertices integral, one can find a* t *so large that the complex admits a linear simplicial* $(\frac{1}{2})^t - \mathbb{Z}^3$ *primitive subdivision.*

Proof. Because of the convexity of its cells, such a complex can be subdivided simplicially into $\mathbb{Z}^3$-elementary simplices. It suffices thus to stipulate an iterative process of elementarily subdividing these simplices (with boundary always to be subdivided in the standard way) which will eventually give, for some t depending on p and q, a $(\frac{1}{2})^t - \mathbb{Z}^3$ primitive subdivision (simplices already primitive will always be further subdivided as in Figure 9).

We assume $X_1, Y_{q-1}, X_{q-1}, Y_1$ are distinct, for otherwise, $p = 1$ or $q - 1$ and we can attain primitivity in just one step. We join these four points in cyclic order, and then further subdivide this "rhombus" — see Figure 10 — by half-integral points, if any, which lie on the interiors of its edges. The subdivided edge path $X_1Y_{q-1}X_{q-1}$ will be our $\mathcal{C}$, and the subdivided edge path $Y_1X_1Y_{q-1}$ our $\mathcal{D}$, so $\mathcal{C} \cap \mathcal{D}$, i.e., the subdivided edge X_1Y_{q-1}, is a subcomplex of both. Using Lemma 1 we choose elementary subdivisions of the components of the complement of the "rhombus" and extend this to a subdivision $\mathbf{K}_{\mathcal{C}}(p,q) \cup \mathbf{K}_{\mathcal{D}}(p,q)$.

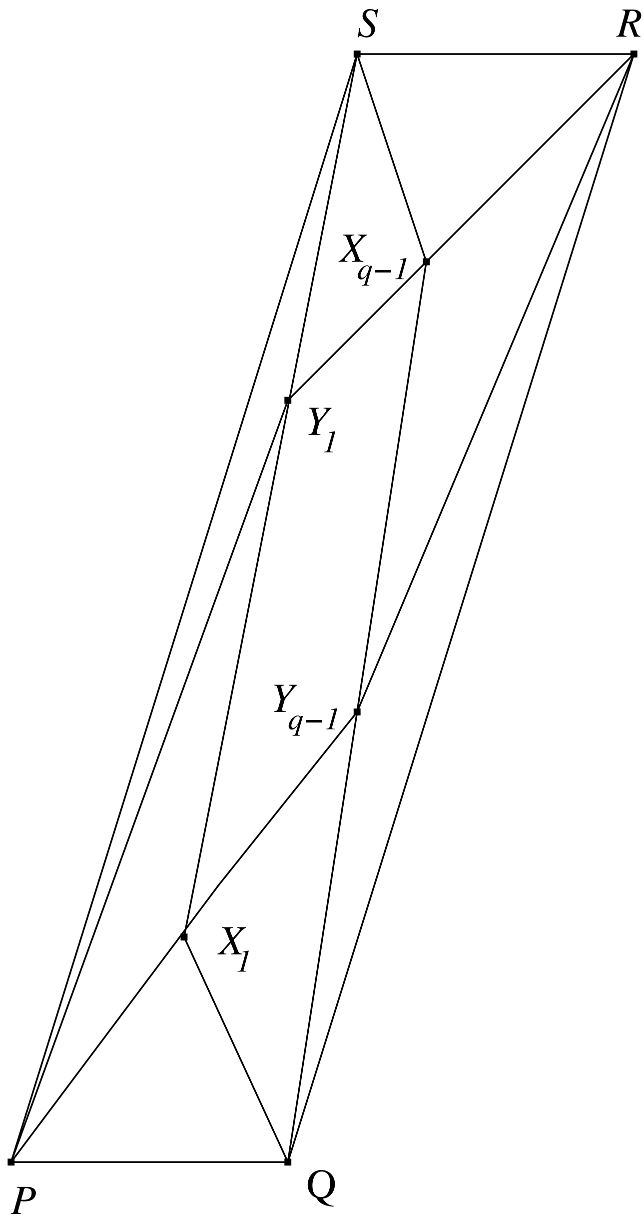


Figure 10.

As X_1 and X_{q-1}, respectively Y_1 and Y_{q-1}, lie on the first and $(q-1)$th parallels to PQ, respectively SP, using Lemma 2 and the symmetry ϕ we see that the new "q", of any tetrahedron of this subdivision, is less than $q-2$. So all tetrahedra eventually become primitive. $\square$

We show next that, without convexity, the first step of the last proof - i.e., the 3-dimensional analogue of Lemma 1 — is false.

A NON-CONVEX 3-CELL. *There is a region Ω of $\mathbb{R}^3$, whose boundary is uniquely triangulable as a linearly embedded 6-vertex simplicial 2-sphere, which meets $\mathbb{Z}^3$ only in these six vertices, and which cannot be simplicially subdivided into elementary tetrahedra.*

Two of the eight boundary triangles, abc and ABC, of this non-convex 3-cell, will lie, respectively, on the horizontal planes $z = 0$ and $z = 1$, while the remaining six, $\{abB, aAB\}, \{bcC, bBC\}$ and $\{caA, cCA\}$, will form three sloping quadrilaterals, $abAB, bcBC$ and $caCA$, all *flexed inwards* along the diagonals aB, bC and cA. In other words, these three diagonals are on the boundary of Ω, but the remaining three diagonals, Ab, Bc and Ca, are all outside Ω. This rules out an elementary simplicial subdivision, because no tetrahedron of such a subdivision could be incident to the face ABC of Ω.

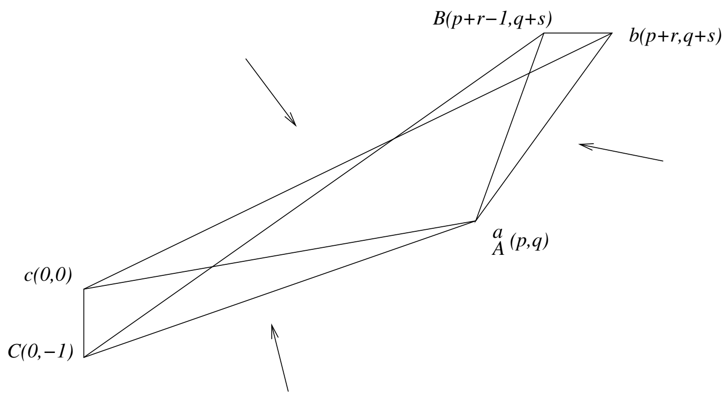


Figure 11.

To arrange this, let the (x, y) projections of the vertices be as in Figure 11, which thus is Ω viewed from the top. More precisely, let $c = (0, 0, 0)$, $C = (0, -1, 1)$, $a = (p, q, 0)$, $A = (p, q, 1)$, $b = (p + r, q + s, 0)$, $B = (p + r - 1, q + s, 1)$, for any four positive integers p, q, r, s, satisfying $\begin{vmatrix} p & q \\ r & s \end{vmatrix} = 1$ and $r = q + 1$ (e.g., $p = 7$, $q = 4$, $r = 5$, $s = 3$). The primitivity of ABC follows from $\begin{vmatrix} p & q+1 \\ r-1 & s \end{vmatrix} = \begin{vmatrix} p & r \\ q & s \end{vmatrix} = 1$ and these equations also ensure that ab slopes more than ca, and AB more than CA. That the quadrilaterals are flexed inwards on the asserted diagonals follows by viewing them from outside along the indicated arrows.

Note that *the above example has the least number of vertices.* A simplicial 4-vertex 2-sphere is the boundary of a tetrahedron, and a 5-vertex simplicial 2-sphere either the boundary of 2 tetrahedra on a common triangle, or 3 tetrahedra around a common edge.

Remark 6. Nevertheless, we emphasize that *Proposition* 6 *itself is true even if the cells of the complex are not convex.* To see this note that all planes, which pass through three affinely independent vertices of the complex, give us a *canonical subdivision* into smaller convex cells with vertices in a bigger lattice $\frac{1}{l}\mathbb{Z}^3$. We choose a simplicial subdivision of this canonical subdivision with vertices in the same lattice $\frac{1}{l}\mathbb{Z}^3$. If l contains primes other than 2, we now move the new vertices slightly to get a combinatorially isomorphic linear simplicial subdivision with vertices in some $(\frac{1}{2})^t - \mathbb{Z}^3$, and then proceed to a primitive subdivision just as before.

To choose more efficient chain pairs $(\mathcal{C}, \mathcal{D})$, we need some more facts regarding the disposition of the $q-1$ points of the lattice $\frac{1}{2}\mathbb{Z}^3$ within the central parallelogram.

LATITUDES AND LONGITUDES. The lattice point X_2 lies either on QX_1 produced (this happens iff $q/p > 2$) or on PX_1 produced; accordingly we will refer to the straight portions of the maximal X-chain parallel to QX_1 or PX_1 as latitudes. Likewise, Y_2 lies on PY_1 produced or on SY_1 produced, and accordingly, straight portions of the maximal Y-chain parallel to PY_1 or SY_1 will be called longitudes. Their slopes can be easily computed by using the fact that (besides $z = \frac{1}{2}$) the coordinates of the end points are $X_1 = \frac{1}{2}(1,1)$, $X_{q-1} = \frac{1}{2}(p, q-1)$, $Y_1 = \frac{1}{2}(r,s)$ and $Y_{q-1} = \frac{1}{2}(p+1-r, q-s)$, where $\begin{vmatrix} r & s \\ p & q \end{vmatrix} = 1$ (because PY_1S and $RY_{q-1}S$ are congruent elementary triangles). Note that $1 \le r < p$ represents the element q^{-1} of $(\mathbb{Z}/p)^\times$, and $1 \le s < q$ is given by $s = -p^{-1} \in (\mathbb{Z}/q)^\times$.

This is the number s of Lemma 3 below, a known[4] classification of the standard models. Then Lemma 4 gives a geometrical reformulation of Lemma 2 and its Y-analogue in terms of (latitudes, horizontals) and (longitudes, verticals). Here, $\operatorname{lat}(IJ) = 0$ if the latitudes through I and J coincide, and otherwise $\operatorname{lat}(IJ)$ is one more than the number of intervening latitudes of the other three nonnegative numbers $\operatorname{hor}(IJ)$, $\operatorname{long}(IJ)$ and $\operatorname{ver}(IJ)$ are defined analogously.

Lemma 3. $\mathbf{T}(p,q)$ *is isomorphic to* $\mathbf{T}(p',q')$ *iff* $q' = q$ *and* p' *is one of the four numbers* $\{p, q-p, s, q-s\}$.

Lemma 4. *If* IJ *is an* $\frac{1}{2}\mathbb{Z}^3$*-elementary edge of the parallelogram then* $AUIJ$ *and* $UBIJ$ *are isomorphic to* $\frac{1}{2}\mathbf{T}(\operatorname{lat}(IJ), \operatorname{hor}(IJ))$, *while* $CVIJ$ *and* $VDIJ$ *are isomorphic to* $\frac{1}{2}\mathbf{T}(\operatorname{long}(IJ), \operatorname{ver}(IJ))$.

Proof. For the "if" part of Lemma 3 note that the affine transformation $(x,y,z) \longrightarrow (-x+y-z+1, y, z)$ maps $\mathbb{Z}^3$ onto $\mathbb{Z}^3$ and $\mathbf{T}(p,q)$ onto $\mathbf{T}(q-p,q)$.

[4]Lemma 3 is a particular case of Theorem 5.6 of Reznick [**Rez**]; in [**Ka**], Proposition 8, two of the 4 possibilities were overlooked.

Again, if we use the unimodular matrix $\begin{bmatrix} q-s & -p+r \\ q & -p \end{bmatrix}$ to transform each plane $z =$ constant, and then reflect in the central plane $z = \frac{1}{2}$, we get the $\mathbb{Z}^3$-preserving affine transformation $\Psi(x, y, z) = (x(q-s) + y(-p+r), xq - yp, 1-z)$, which maps $\mathbf{T}(p, q)$ onto $\mathbf{T}(q-s, q)$.

For the converse note that an affine transformation Ψ will map $\mathbf{T}(p, q)$ onto $\mathbf{T}(p', q')$ iff $(\Psi(A), \Psi(B), \Psi(C), \Psi(D))$ is one of the $4! = 24$ permutations of the vertices (A', B', C', D') of the second tetrahedra. Here we use primes to denote corresponding points of second tetrahedron, so $A' = A$, $B' = B$, $C' = C$ and $D' = (p', q', 1)$. we have $q' = q$ because the two tetrahedra have the same volume; so Ψ is volume preserving, and $\Psi(\mathbb{Z}^3) = \mathbb{Z}^3$ iff $(0, 1, 0) = (p/q + 1)A - (p/q)B - (1/q)C + (1/q)D$ is mapped to an integral point $\Psi(0, 1, 0) = (p/q + 1)\Psi(A) - (p/q)\Psi(B) - (1/q)\Psi(C) + (1/q)\Psi(D)$. Checking this for each of the 24 cases (cf. Reznick [**Rez**]) one sees that such a Ψ exists only if $p' = p$, $q - p$, s or $q - s$.

For the isomorphism $\Psi : \mathbf{T}(p, q) \simeq \mathbf{T}(q-s, q)$ defined in the first paragraph $(\Psi(A), \Psi(B), \Psi(C), \Psi(D)) = (C', D', A', B')$ so Ψ maps (P, Q, R, S, U, V) in order onto (P', S', R', Q', V', U'). From this it follows that it maps the verticals, resp. longitudes, of $\mathbf{T}(p, q)$ onto the horizontals, resp. latitudes, of $\mathbf{T}(q-s, q)$, with the $q-1$ points Y_k of the central section of $\mathbf{T}(p, q)$ mapping in order on the $q-1$ points X'_k of $\mathbf{T}(q-s, q)$.

This last observation shows that it is enough to prove only the first part of Lemma 4. For this note that, by Lemma 2, both $AUIJ$ and $UBIJ$ are isomorphic to $\frac{1}{2}\mathbf{T}(j_1 - i_1, j_2 - i_2)$, but $\mathrm{hor}\,(IJ) = j_2 - i_2$, and $\mathrm{lat}\,(IJ)$ is either the same as $j_1 - i_1$ (this when $q > 2p$ for then the latitudes have equations $x =$constant) or equal to $(j_2 - i_2) - (j_1 - i_1)$ (this when $q < p$ and the latitudes have equations $y - x =$constant). $\square$

For somewhat higher values of (p, q), than those used in Figures 5 and 6, it is helpful to plot the above four kinds of lines using the following device, because it avoids the congestion of longitudes and verticals, and gives a more balanced picture.

SQUARING. Our problem belongs to three dimensional affine geometry, i.e., all that we are saying about $\mathbf{T}$ and $\mathbb{Z}^3$ applies equally well to $\mu(\mathbf{T})$ and $\mu(\mathbb{Z}^3)$, where μ is any affine transformation of 3-space. For instance, by applying the linear $\mu : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ having matrix

$$\begin{bmatrix} q & -p & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

we obtain a $\mu\mathbf{T}$ having edges AB and CD of the same length, and its section $z = \frac{1}{2}$ is a square, the new or "squared" x- and y-coordinates of the various points being $\frac{1}{2}$ times the numbers indicated in Figure 12.

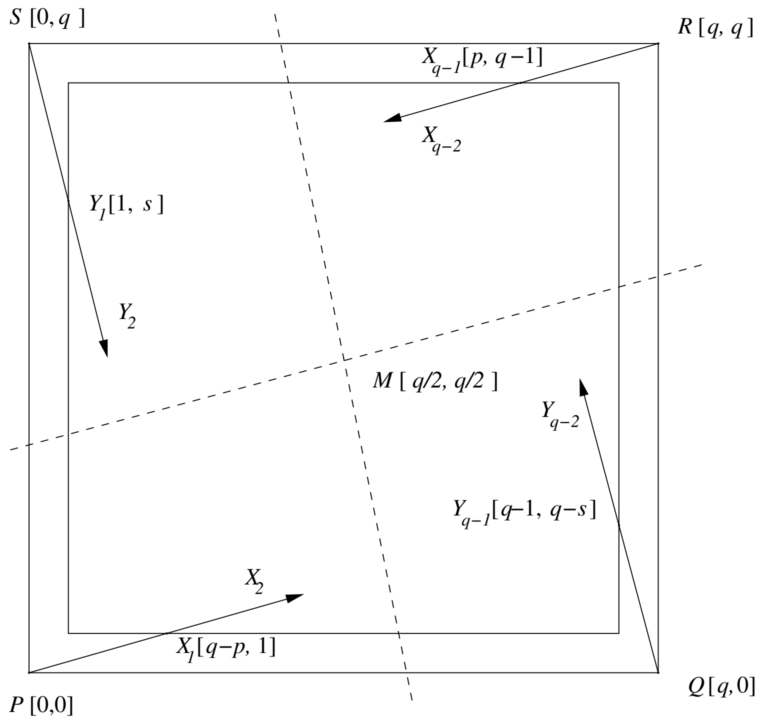


Figure 12.

Note that the $q-1$ equally spaced verticals, being parallel to SP, are now in fact "vertical", and the spacing between them is the same as between the $q-1$ horizontals. There is one and only one point of the lattice $\mu(\mathbb{Z}^3)$ on each horizontal and vertical, and these points can be found by starting from any end point and making some obvious MOVES. For example, we start from X_1 in the direction PX_1 — because $PX_1 < QX_1$ in Figure 12, in case $PX_1 > QX_1$ start in the direction QX_1 — till we meet the second horizontal in X_2, and so on, as long as we stay in the square. Having thus obtained the first latitude $X_1X_2 \ldots X_i$, we then make one move from X_i parallel to QX_1 to find X_{i+1} on the next horizontal, after which we again make moves parallel to PX_1 and obtain the second latitude, etc. In some figures below, we will only draw the smaller square passing through the four end points. The "crossword" symmetry of these figures reflects that $A \leftrightarrow B$, $C \leftrightarrow D$ is an automorphism of $\mathbf{T}(p,q)$, and the fact that the end points $X_{q-1}, X_1, Y_1, Y_{q-1}$ have, respectively, squared coordinates $(p, q-1), (q-p, 1), (1, s), (q-1, q-s)$ is related to Lemma 3: The symmetries of the square $PQRS$ induce the four isomorphisms $\mathbf{T}(p,q) \simeq \mathbf{T}(p',q)$. Note that if $p = s$ or $q-s$, then one has only two numbers, and so only two isomorphic tetrahedra; but nevertheless, Lemma 3 shows that one may always assume $p < q/2$, as we will do in the next proposition.

ALMOST MAXIMAL CHAINS. An X-chain $\mathcal{C} = \overline{X}_1\overline{X}_2\dots\overline{X}_r$ is called almost maximal if its edges $\overline{X}_i\overline{X}_{i+1}$ join vertices on the same or adjacent latitudes, i.e., iff lat $(\overline{X}_i\overline{X}_{i+1}) = 0$ or 1. Thus, besides $\mathcal{X}$, there are many other almost maximal chains; for any such $\mathcal{C}$, Lemma 4 tells us that the $\frac{1}{2}\mathbb{Z}^3$-elementary tetrahedra $\overline{X}_i\overline{X}_{i+1} * AU$ and $\overline{X}_i\overline{X}_{i+1} * UB$ are either $\frac{1}{2}\mathbb{Z}^3$-primitive or $\frac{1}{2}\mathbb{Z}^3$-ALMOST PRIMITIVE, i.e., isomorphic to $\frac{1}{2}\mathbf{T}(1,t)$ for some t. Almost maximal Y-chains are defined analogously. So, *if* $(\mathcal{C},\mathcal{D})$ *is a pair of almost maximal chains, with* $\mathcal{C}\cap\mathcal{D}$ *a subcomplex, then, by using the "if" part of Theorem* 1 *to further subdivide the* $\frac{1}{2}\mathbb{Z}^3$*-elementary subdivision* $\mathbf{K}_{\mathcal{C},\mathcal{D}}(p,q)$*, we obtain a* $\frac{1}{4}\mathbb{Z}^3$*-primitive subdivision of* $\mathbf{T}(p,q)$. As a warm-up for the proof of Theorem 2, we first show that often one of these chains $(\mathcal{C},\mathcal{D})$ can be chosen arbitrarily.

Proposition 7. *The two X-chains, which run from X_1 to X_{q-1} on the boundary of the convex hull, of the $q-1$ points of $\frac{1}{2}\mathbb{Z}^3$ lying in the interior of the section $z=\frac{1}{2}$ of $\mathbf{T}(p,q)$, $p<q/2$, are almost maximal if and only if p divides $q-1$ or $q+1$.*

For any such (p,q), we obtain a $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision of $\mathbf{T}(p,q)$ by subdividing $\mathbf{K}_{\mathcal{C},\mathcal{D}}(p,q)$, where $\mathcal{C}$ is one of these two chains, and $\mathcal{D}$ is any almost maximal Y-chain. This follows because, $\mathcal{C}$ being on their convex hull, no edge path $\mathcal{D}$, having as vertices some or all of the $q-1$ lattice points can cross it. Thus $\mathcal{C}\cap\mathcal{D}$ is a subcomplex for any $\mathcal{D}$. If $\mathcal{D}$ is maximal, the top half of $\mathbf{K}_{\mathcal{C},\mathcal{D}}(p,q)$ is already $\frac{1}{2}\mathbb{Z}^3$-primitive, and the bigger lattice $\frac{1}{4}\mathbb{Z}^3$ is used only to make the bottom primitive.

Proof. We assert that, *for $q/p>2$, the maximal X-chain has p latitudes, the first having $t=[q/p]$ vertices, and all others having either t or else $t+1$ vertices.* This centrally symmetric partition into p parts of size t or $t+1$ will be called the X-PARTITION of $q-1$.

To see this note that in this case the latitudes are configured as for the (3,8) picturized in Figure 5, i.e., they lie on p equally spaced "verticals" (not to be confused with *verticals* (!) which are parallel to PS). The first of these "verticals" passes through Q, and the last through S. Since PS has slope q/p, the first has exactly $t+1$, $t=[q/p]$, points of the lattice $\frac{1}{2}\mathbb{Z}^3$, of which one, Q, is on the boundary.[5] Since the parallelogram intercepts the same length of each "vertical" it follows that all have either t or $t+1$ points of $\frac{1}{2}\mathbb{Z}^3$.

Next, consider the X-GON $\mathcal{P}_X$, i.e., the undirected simple closed edge path consisting of the first and the last latitude, and the two edge paths obtained by joining the initial, respectively final, vertex of each latitude, to the initial, respectively final, vertex of the next latitude. Thus the two

[5]For the same reason, the ith "vertical" has $[iq/p]-[(i-1)q/p]$ points.

chains $\mathcal{C}$ which run on $\mathcal{P}_X$ from X_1 to X_{q-1} are both almost maximal, and each of the $q-1$ points X_i is either a vertex of $\mathcal{P}_X$, or is enclosed within it.

So, it only remains to check that $\mathcal{P}_X$ *is convex iff* p *divides* $q-1$ *or* $q+1$. Convexity is equivalent to saying that the internal angle at each vertex v of $\mathcal{P}_X$ is at most $180°$. By central symmetry we need to consider only the initial vertices v of the latitudes. At the initial vertices of the first and pth latitudes the angle is clearly less than $180°$. However, at an initial vertex v of any other latitude the internal angle of $\mathcal{P}_X$ exceeds $180°$ iff the preceding latitude has more vertices (i.e., iff, of the edges of $\mathcal{P}_X$ meeting in v, the one on the left has slope, $t+1$, one more than the one on the right). Thus, either all the parts of the X-partition have the same size t, or only the initial and final parts are of size t, and all others are of size $t+1$. So, either $q-1 = pt$, or $q-1 = 2t + (p-2)(t+1)$, i.e., $q+1 = p(t+1)$. $\square$

Remark 7. We thus have $\frac{1}{4}\mathbb{Z}^3$-primitive subdivisions of many elementary tetrahedra, e.g., all those isomorphic to a $\mathbf{T}(p,q)$ with $p \leq 4$, and, using Lemma 3, this covers all having $q \leq 11$. The case $(p,q) = (5,12)$ is the smallest which escapes Proposition 7, and now, not only is the X-gon $\mathcal{P}_X$ non-convex, its intersection with the maximal Y-chain — shown dotted in Figure 13a — is not a subcomplex. So, if $\mathcal{C}$ is on $\mathcal{P}_X$, the almost maximal Y-chain $\mathcal{D}$ cannot be arbitrary. But, *in an amazingly large number of cases, including all with* $q \leq 20$*, the intersection* $\mathcal{P}_X \cap \mathcal{P}_Y$ *of the* X*-gon and the* Y*-gon is a subcomplex*, e.g., as Figure 13b shows, if $(p,q) = (7,19)$, then $\mathcal{P}_X \cap \mathcal{P}_Y$ consists of 6 edges and their 10 vertices. In all these cases, we obtain a $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision by choosing the complementary $\mathcal{D}$ on $\mathcal{P}_Y$. *However,* $\mathcal{P}_X \cap \mathcal{P}_Y$ *is not always a subcomplex:* Figure 14 shows that for the case $(p,q) = (12,31)$, $\mathcal{P}_X$ and $\mathcal{P}_Y$ have no common edge, but $\mathcal{P}_X \cap \mathcal{P}_Y$ contains, besides 22 vertices, 4 other isolated points. Now, for a pair $(\mathcal{C},\mathcal{D})$ with $\mathcal{C} \subset \mathcal{P}_X$ and $\mathcal{D} \subset \mathcal{P}_Y$, $\mathcal{C} \cap \mathcal{D}$ is not a subcomplex (but, using a small detour in one of the 2 chains, one again gets a pair of almost maximal chains with intersection a subcomplex).

We do not need Proposition 7, or any of the constructions just mentioned, in the proof of Theorem 2. However they are of independent interest, and are what first convinced us about the truth of this theorem. In all these constructions, we stayed away from the barycenter M, thinking that this was needed to avoid bad intersections. However, somewhat surprisingly, "moving towards M" is exactly what gave us a general construction of the required chain pairs.

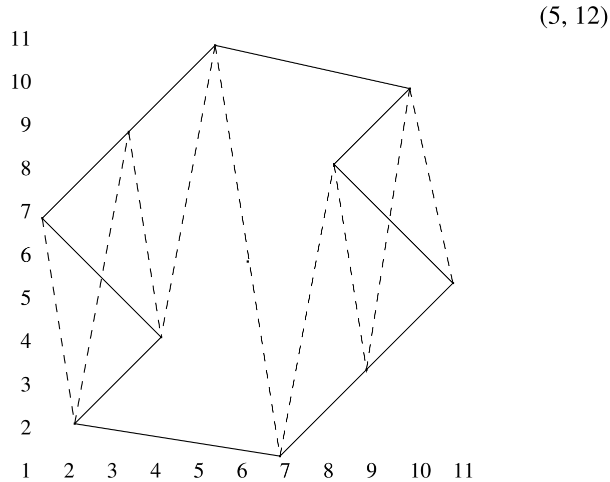


Figure 13a.

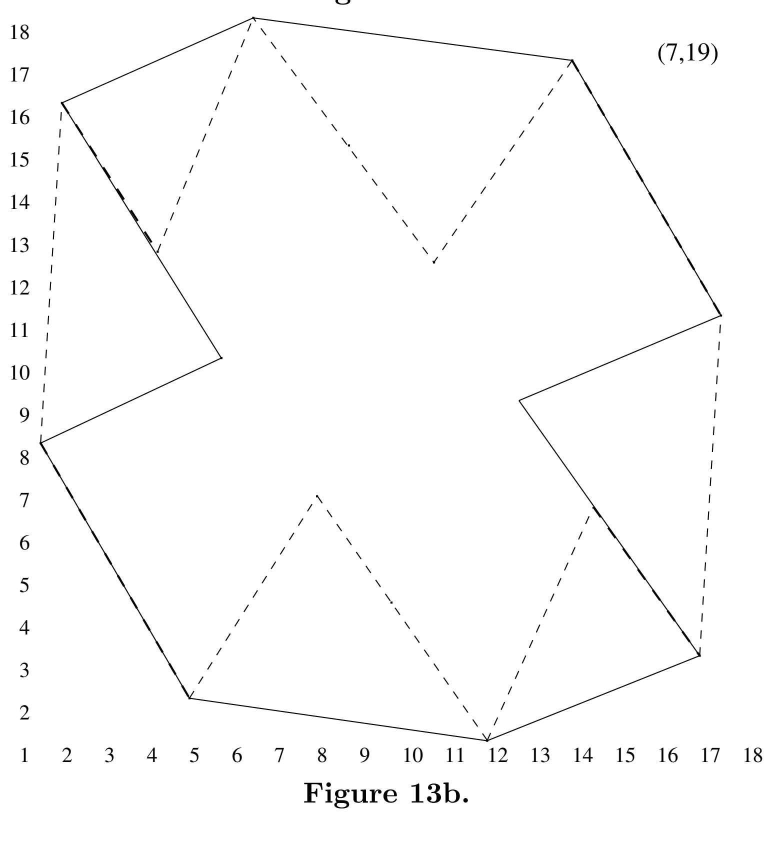


Figure 13b.

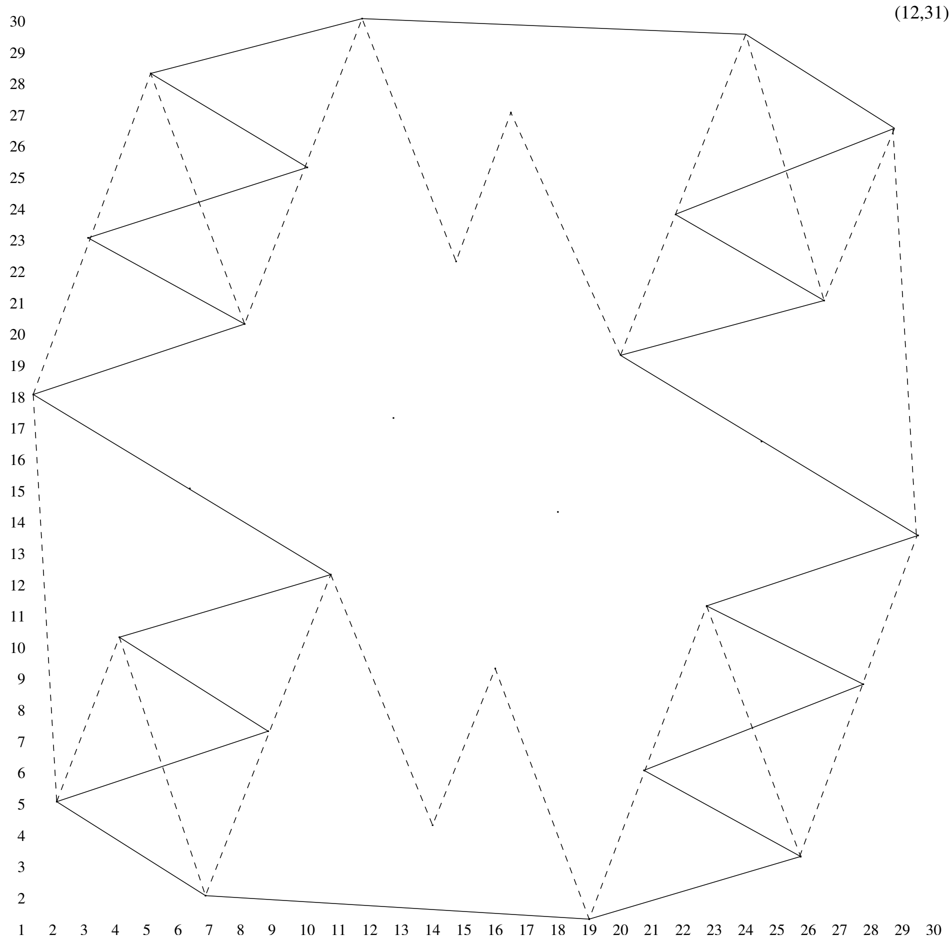


Figure 14.

QUADRANTS. Choosing M as "origin" we draw — see Figure 12 — through M two "axes" in the directions parallel to the latitudes and the longitudes. If q is even, $M \in \frac{1}{2}\mathbb{Z}^3$, and each "axis" is itself a latitude or longitude; however, for q odd, when $M \notin \frac{1}{2}\mathbb{Z}^3$, it may or may not be one. We adopt the usual terminology for the four parts of the parallelogram, e.g., the "first quadrant" is that which contains R. The idea now is to (essentially) confine $\mathcal{C}$ and $\mathcal{D}$ to distinct pairs of opposite quadrants: Since then $M \in \mathcal{C} \cap \mathcal{D}$, this idea works best only for q even, so this is the case which we will treat first.

Proposition 8. *For all q even, there is a pair $(\mathcal{C}, \mathcal{D})$ of centrally symmetric almost maximal chains, with $\mathcal{C} \cap \mathcal{D} = \{M\}$.*

So, by the argument as before, all $\mathbf{T}(p,q)$ with q even must have a $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision. We note that in the following we will argue from the squared representation of the central section.

Proof. First, suppose that the slopes of the "axes" have opposite signs, say, the latitudes have positive slope and the longitudes negative (the opposite case is similar). This is *the case* $q-p < q/2 < s$, and is the one that was shown in Figure 12. Note that the open third quadrant (i.e., that containing P) intercepts a length bigger than PX_1 from all latitudes, starting with the first, and before that through M. So we can choose, on each of these intercepts, one point of the lattice $\mu(\mathbb{Z}^3)$. Using these chosen points, the barycenter M, and the symmetric points of the first quadrant (i.e., that containing R), we obtain a vertex subsequence of the maximal X-chain $X_1X_2\ldots X_{q-1}$, which determines an almost maximal X-chain $\mathcal{C}$, which but for M, is contained in the interior of the third and first quadrants. Likewise, because the longitudinal intercepts contained in the open second quadrant have length bigger than SY_1, we can define a centrally symmetric almost maximal Y-chain $\mathcal{D}$, which, but for M, is contained in the interior of the second and fourth quadrants. So $\mathcal{C}\cap\mathcal{D} = \{M\}$.

Now suppose that both latitudes and longitudes have positive slope (the opposite case when both slopes are negative is similar), i.e., that *both* $q-p$ *and* s *are less than* $q/2$. As shown in Figure 15, now X_1 and Y_1 may both be in the third quadrant, or one of them may be instead in the neighbouring, i.e., fourth or second respectively, quadrant, but we assert that *one cannot have* X_1 *in the fourth quadrant and* Y_1 *in the second quadrant*. For this to happen, the slope of MX_1, i.e., $(q/2-1)\div(q/2-(q-p))$, must exceed the slope s of a longitude, and the slope of MY_1, i.e., $(q/2-s)\div(q/2-1)$, must be less than the slope $1/(q-p)$ of a latitude. So $q/2-1$ has to be bigger than both $s(q/2-(q-p))$ and $(q/2-s)(q-p)$. If $s < q-p$ the second number is bigger, if $s = q-p$ both are equal, and if $s > q-p$ the first number is bigger. In any case the bigger of the numbers exceeds the minimum $(q/4)^2$ of $x(q/2-x)$, $0 \le x \le q/2$. So $q/2-1 > q^2/16$, i.e., $(q-4)^2 < 0$, which is absurd.

If say Y_1 is in the second quadrant, then, noting that the longitudinal intercept contained in this quadrant have a length bigger than PY_1, we choose as above an almost maximal and symmetric $\mathcal{D}$ through the interiors of the second and fourth quadrants, except for the one point M. Also, the latitudinal intercepts in the third quadrant being of length at least PX_1, there is an analogous $\mathcal{C}$ through the third and first quadrant. The case when X_1 is in the fourth and Y_1 in the third quadrant, is exactly similar, only the quadrant pairs of the chains are now switched. So, in all these cases, $\mathcal{C}\cap\mathcal{D} = \{M\}$.

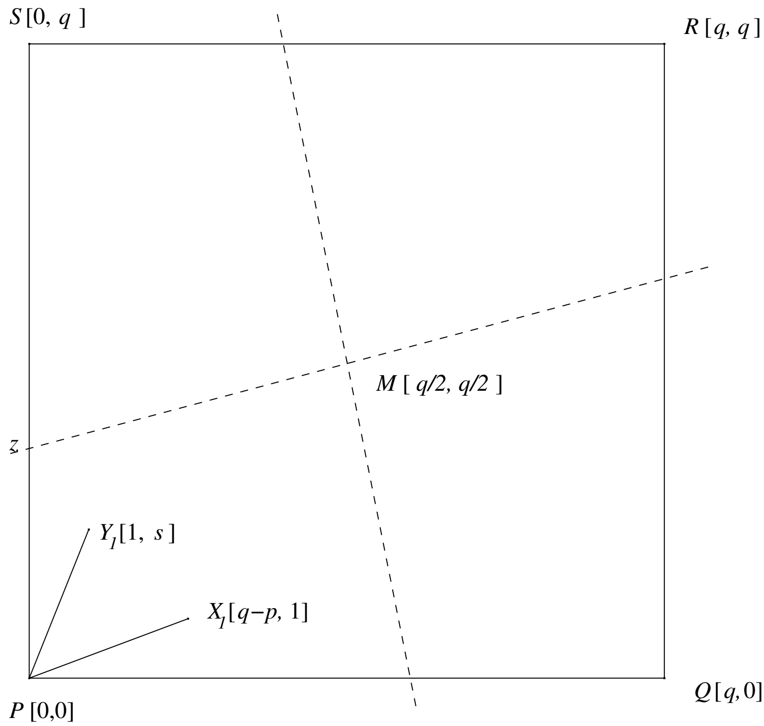


Figure 15.

There remains only the case of Figure 15, when of course $\mathcal{C}$ and $\mathcal{D}$ cannot be completely (but for M) in distinct pairs of opposite quadrants. Now choose one of the chains, say $\mathcal{D}$, just as before, so it runs through the third and first quadrants. Moreover, since the portion of the third quadrant lying above PX_1 produced also has longitudinal intercepts of length bigger than PY_1, we can arrange that $\mathcal{D}$ is strictly above the first latitude, and so also strictly below the last latitude. For $\mathcal{C}$, we choose an "initial portion" on the first latitude till we are in the fourth quadrant. The "middle portion" of $\mathcal{C}$ runs between this and the symmetric "final portion", through the fourth and second quadrants, and is defined just as before using the fact that the latitudinal intercepts in the fourth quadrant have length exceeding PX_1. This gives again a symmetric almost maximal pair with $\mathcal{C} \cap \mathcal{D} = \{M\}$. $\square$

To finish the proof of Theorem 2 we will use, firstly, the fact that an analogous conclusion is valuable even if q is odd.

Lemma 5. *Any $\mathbf{T}(p, q)$ having a pair $(\mathcal{C}, \mathcal{D})$ of almost maximal chains with $\mathcal{C} \cap \mathcal{D} = \{M\}$ admits a $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision.*

More generally, a similar construction shows that the same is true if $\mathcal{C}' \cap \mathcal{D}'$ is a subcomplex of $\mathcal{C}'$ and $\mathcal{D}'$, the subdivisions of $\mathcal{C}$ and $\mathcal{D}$ obtained by considering middle points of edges as new vertices.

Proof. For the remaining *case q odd*, we will describe a modification of the previous construction, which only uses the fact that $M \in \frac{1}{4}\mathbb{Z}^3$. We begin by constructing the $\frac{1}{2}\mathbb{Z}^3$-elementary subdivisions $\mathbf{K}_\mathcal{C}(p,q)$ and $\mathbf{K}_\mathcal{D}(p,q)$, of the bottom and the top half, just as before, but take the following extra care in choosing the sub-triangulations $\mathbf{P}_\mathcal{C}(p,q)$ and $\mathbf{P}_\mathcal{D}(p,q)$, of the parallelogram $\mathbf{P}$, used in defining them.

Let $X\overline{X}$ and $Y\overline{Y}$ denote, respectively, the edges of $\mathcal{C}$ and $\mathcal{D}$ of which M is the middle point. We choose a $\frac{1}{2}\mathbb{Z}^3$-elementary centrally symmetric polygonal region $\mathbf{R} \subseteq \mathbf{P}$, such that four of its vertices are $X, \overline{Y}, \overline{X}$, and Y. Note that this $\mathbf{R}$ can have other vertex pairs $\{E, \overline{E}\}$, and can be non-convex, we only demand that if a segment having M as its middle point has one end point in $\mathbf{R}$, then the entire segment be in $\mathbf{R}$. It is easily seen that such a polygon $\mathbf{R}$ can always be found. The extra care we now take is that the boundary of $\mathbf{R}$ be a subcomplex of both $\mathbf{P}_\mathcal{C}(p,q) \subset \mathbf{K}_\mathcal{C}(p,q)$ and $\mathbf{P}_\mathcal{D}(p,q) \subset \mathbf{K}_\mathcal{D}(p,q)$, and that $\mathbf{P}_\mathcal{C}(p,q)$ and $\mathbf{P}_\mathcal{D}(p,q)$ differ only on the interior of the polygonal region $\mathbf{R}$.

The triangulations $\mathbf{R}_\mathcal{C}(p,q)$ and $\mathbf{R}_\mathcal{D}(p,q)$ of this region *have* to be different, because $X\overline{X}$ is in the former, and $Y\overline{Y}$ in the latter. We nevertheless go ahead and subdivide $\mathbf{K}_\mathcal{C}(p,q)$ just as before to obtain a $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision $\mathbf{K}'_\mathcal{C}(p,q)$ of the bottom half. Recall that in this step the tetrahedra of the second kind are subdivided using Theorem 1, and the remaining tetrahedra, all primitive, are subdivided further by the method of Figure 9. The edge $X\overline{X}$ separates the interior of $\mathbf{R}$ into two parts, one contained in the left polygonal region of $\mathcal{C}$, and the other in the right polygonal region. Removing all simplices of $\mathbf{K}'_\mathcal{C}(p,q)$ incident to the left part creates a TROUGH — see Figure 16 — whose bottom carries a triangulation similar to the left half of $\mathbf{R}_\mathcal{C}(p,q)$, and all walls of this trough are triangulated in the standard way. In particular, the triangulation of the wall of $X\overline{X}$ is a cone over M. So we can refill this trough by coning the remaining walls and bottom over M. We similarly create and refill a trough on the right side. This gives us a $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision $\mathbf{K}''_\mathcal{C}(p,q)$ of the bottom half, which matches with the similarly modified $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision $\mathbf{K}''_\mathcal{D}(p,q)$ of the top half, because, in both cases, the polygonal region $\mathbf{R}$ has now been retriangulated as the cone of its boundary over M. $\square$

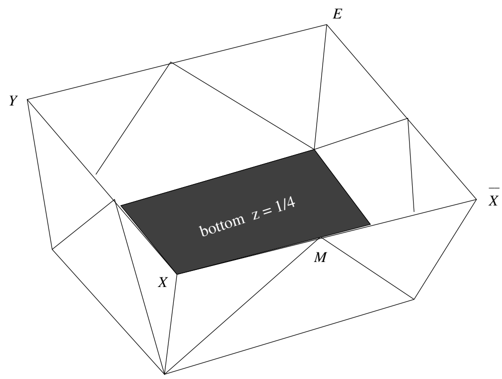


Figure 16.

Another point which we will use in the following is that we can employ chains slightly more general than almost maximal chains, namely, we can often SKIP a latitude (or longitude) provided it contains only two lattice points. For instance, if $q/p > 2$, and X_j is the last vertex of the previous latitude, then the first vertex of the next latitude is X_{j+3}, and using Lemma 2 we know that the join $X_j X_{j+3} * (AU \text{ or } UB)$ is isomorphic to $\frac{1}{2}\mathbf{T}(2,3)$, and so admits a $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision.

Conclusion of proof of Theorem 2. For the remaining case q odd, the argument of Proposition 8 applies verbatim, to yield a centrally symmetric almost maximal pair $(\mathcal{C}, \mathcal{D})$ with $\mathcal{C} \cap \mathcal{D} = \{M\}$, provided the numbers of latitudes and longitudes are both even (e.g., if $q - p$ and s are even and less than $q/2$). Using Lemma 5 we obtain the subdivision.

In all other cases, at least one of the "axes" has lattice points, the lattice points nearest to M on such an "axis" being $\{X_t, X_{t+1}\}$ or $\{Y_t, Y_{t+1}\}$, where $t = (q-1)/2$. We still need to make only a trifling change in the construction of $\mathcal{C}$ or $\mathcal{D}$, as described in the proof of Proposition 8, provided the slopes of the two "axes" have different signs. The trifling change being that we choose *both* $\{X_t, X_{t+1}\}$ or $\{Y_t, Y_{t+1}\}$ from this central latitude or longitude. This again gives a centrally symmetric almost maximal pair $(\mathcal{C}, \mathcal{D})$ with $\mathcal{C} \cap \mathcal{D} = \{M\}$.

This is however not permissible, for one of the two chains, when the slopes of the two "axes" have the same sign, say, positive as in Figure 15 (case negative is similar). Now, in the construction used to prove Proposition 8, one of the two chains, say $\mathcal{C}$, had either not started from the third quadrant, or else had been diverted from it, by means of an "initial portion" on the first latitude, into the neighbouring fourth quadrant, and then $\mathcal{C}$ had remained, till M, in this quadrant. This entails choosing X_{t+1} *before* X_t, which is not allowed.

This difficulty is easily overcome if $\{X_t, X_{t+1}\}$ happen to be the only lattice points of the central latitude. Now, while defining $\mathcal{C}$, we choose the last vertex X_{t-1} of the previous latitude, which is in the fourth quadrant, then skip both $\{X_t, X_{t+1}\}$, and move on to the symmetric first vertex X_{t+2} of the next latitude, which is in the second quadrant. Once again we obtain a pair $(\mathcal{C}, \mathcal{D})$ of centrally symmetric chains with $\mathcal{C} \cap \mathcal{D} = \{M\}$, and, using the remark made above, one can still use these chains as before to make a $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision. In case it is only the central longitude which has two lattice points, and one of the two chains needs to be diverted, we take care to divert $\mathcal{D}$, not $\mathcal{C}$.

If the number of lattice points on each "axis" is more than two, i.e., four or more, we give up central symmetry, and overcome the above difficulty by SHIFTING THE ORIGIN from M to the lattice point X_t (or Y_t) the new "quadrants" being now those which are determined by the latitude and longitude through X_t. Because there are at least 4 lattice points on the central latitude, the distance from X_t to Z, the point where this latitude intersects PS, is bigger than PX_1, thus guaranteeing that all latitudinal intercepts contained in the (now smaller) second quadrant still have length bigger than PX_1. Thus $\mathcal{C}$ can be defined like before: An initial diversion on the first latitude into the fourth quadrant, then a lattice point from each latitude in this (now bigger) quadrant till the point X_t, after which we can continue like this in the third quadrant, thanks to the observation just made, till we meet the final latitude, along which we finally go to X_{q-1}. The other chain $\mathcal{D}$ is in the new third and first quadrants, and definable just as before, because the longitudinal intercepts contained in them, and between the first and final latitudes, have lengths bigger than PY_1. This gives a (non-symmetric) pair of almost maximal chains $\mathcal{C}$ and $\mathcal{D}$, whose intersection is the subcomplex consisting of the single lattice point X_t. So once again a $\frac{1}{4}\mathbb{Z}^3$-primitive subdivision exists. $\square$

References

[Da] V.I. Danilov, *Birational geometry of toric* 3*-folds*, Math. USSR Izv., **21** (1983), 269-280, MR 84e:14008, Zbl 0536.14008.

[GKZ] I.M. Gelfand, M.M. Kapranov and A.V. Zelevinsky, *Discriminants, Resultants and Multidimensional Determinants*, Birkhaüser, Basel, 1994, MR 95e:14045, Zbl 0827.14036.

[Hu] J.F.P. Hudson, *Piecewise Linear Topology*, Benjamin, New York, 1969, MR 40 #2094, Zbl 0189.54507.

[Ka] J.-M. Kantor, *Triangulations of integral polytopes and Ehrhart polynommials*, Beitr. Algebra Geom., **39** (1998), 205-218, MR 99b:52024, Zbl 0893.52008.

[Ka2] ———, *On the width of lattice-free simplices*, Compos. Math., **118** (1999), 235-241, MR 2001h:52011, Zbl 0936.52004

[KKMS] G. Kempf, F. Knudsen, D. Mumford and B. Saint Donat, *Toroidal Embeddings* I, Springer Lecture Notes, **339**, 1973, MR 49 #299, Zbl 0271.14017.

[Re] J.E. Reeve, *On the volume of lattice polyhedra*, Proc. Lond. Math. Soc., **7** (1957), 378-395, MR 20 #1954, Zbl 0080.26701.

[Rez] B. Reznick, *Lattice point simplices*, Discrete Math., **60** (1986), 219-242, MR 87i:52022, Zbl 0593.52008.

[Wh] G.K. White, *Lattice tetrahedra*, Can. J. Math., **16** (1964), 389-396, MR 28 #5041, Zbl 0124.02901.

Received March 24, 2002 and revised October 23, 2002.

Institut de Mathematique de Jussieu
175, rue du Chevaleret
75013 Paris
France
E-mail address: kantor@math.jussieu.fr

Department of Mathematics
Panjab University
Chandigarh 160014
India
E-mail address: sarkaria_2000@yahoo.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

CYCLIC PROPERTIES OF VOLTERRA OPERATOR

Fernando León-Saavedra and Antonio Piqueras-Lerena

A bounded linear operator T defined on a Hilbert space H is said to be supercyclic if there exists a vector $x \in H$ such that the set $\{\lambda T^n x \; : \; n \in \mathbb{N}, \lambda \in \mathbb{C}\}$ is dense in H. In the present work, two open questions posed by N. H. Salas and J. Zemánek respectively, are solved. Namely, we will exhibit that the classical Volterra operator V and the identity plus Volterra operator $I + V$ are not supercyclic.

1. Introduction.

This paper deals with the classical Volterra operator V which was introduced in 1896. It is defined on the Hilbert space $L^2[0,1]$ by

$$Vf(x) = \int_0^x f(s)\,ds.$$

An operator T on a Hilbert space H is said to be supercyclic if there exists a vector $x \in H$ such that the projective orbit $\{\lambda T^n x \; : \lambda \in \mathbb{C}\,, n \in \mathbb{N}\}$ is dense in H. The concept of supercyclicity was introduced originally in [**HW**] by Hilden and Wallen. Supercyclicity stands in the midway between hypercyclicity and cyclicity. An operator is said to be hypercyclic if there exists a vector whose orbit under T is dense. On the other hand, if the linear span of some orbit is dense, the operator is called cyclic.

We have two goals:

a) To show that V cannot be supercyclic on $L^2[0,1]$, and
b) the identity plus Volterra operator $I+V$ is not supercyclic on $L^2[0,1]$.

The first question was posed by N. H. Salas in [**Sa**] the second one by J. Zemánek in personal communication. In Section 2 we will renew acquaintance with the Volterra operator by proving that V and $I+V$ are not hypercyclic, however they are cyclic. Section 3 is devoted to prove our main result.

Volterra operator has been studied by several authors. The norm of Volterra operator is $2/\pi$ (see [**Ha**, Problem 149]). The problem's book of P. R. Halmos contains several nice results (some of them not so elementary) related with Volterra operator. The asymptotic behaviour of the norm $\|V^n\|$ is described in [**LR**]. The most interesting fact about the Volterra operator is the determination of its invariant subspace lattice (see [**Co**, Chapter 4], and [**Br**], [**Dix**], [**Don**], [**Ka**] and [**Sar**]). Although Volterra operator is more than a hundred years old however still there exist several open questions, for example, it is not known the exact norm $\|V^n\|$ (see [**LR**]); in [**Ts**] appear new results about Volterra operator.

2. Hypercyclicity and cyclicity. Elementary facts.

The Volterra operator is quasinilpotent. Thus the orbit of every vector converges to zero. Therefore V cannot be hypercyclic.

For the identity plus Volterra case the argument is not so easy. The following result was pointed to the authors by J. Zemánek:

Proposition 2.1. *Identity plus Volterra operator is not Hypercyclic on* $L^2[0,1]$.

Proof. The proof is based in this fact: The inverse of $(I+V)$ is power bounded (see [**Ha**, Problem 150]). Thus the orbit of any vector under $(I+V)^{-1}$ is bounded, therefore $(I+V)^{-1}$ cannot be hypercyclic. The result follows from a result of Herrero and Kitai which asserts that an invertible operator is hypercyclic if and only if its inverse is hypercyclic (see [**HK**]). $\square$

However both operators are cyclic. Basically this fact is consequence of Weierstrass's Theorem.

Proposition 2.2. *Volterra and identity plus Volterra operators are cyclic.*

Proof. Let us denote by $L^2_{\mathbb{R}}[0,1]$ the subspace $\{f \in L^2[0,1]$: such that $f[0,1] \subset \mathbb{R}\}$. The orbit of the identity function 1 under V is the set

$$\mathrm{Orb}(V,1) = \left\{1, x, \frac{x^2}{2}, \dots, \frac{x^n}{n!}, \dots\right\}.$$

By Weierstrass's Theorem, the linear span of $\mathrm{Orb}(V,1)$ is dense in $L^2_{\mathbb{R}}[0,1]$. That is, V is cyclic on $L^2_{\mathbb{R}}[0,1]$. Pick $f \in L^2[0,1]$ and $\varepsilon > 0$. The function $f = u+iv$ with $u, v \in L^2_{\mathbb{R}}[0,1]$, therefore there exists polynomials p_u, p_v such that $\|p_u(V)1 - u\|^2 < \varepsilon/2$ and $\|p_v(V)1 - v\|^2 < \varepsilon$. Thus

$$p_u(x) = u_0 + u_1 x + \cdots + u_n x^m \quad p_v(x) = v_0 + v_1 x + \cdots + v_m x^m$$

with $u_i, v_i \in \mathbb{R}$, let us consider $p(z) = \sum_{k=0}^m a_k z^k$ with $a_k = u_k + iv_k$, $k = 0, \dots, m$, and compute

$$\begin{aligned}\|f - p(V)(1)\|^2 &= \|u + iv - p_u(V)(1) - ip_v(V)(1)\|^2 \\ &= \|u - p_u(V)(1)\|^2 + \|v - p_v(V)(1)\|^2 < \varepsilon,\end{aligned}$$

therefore 1 is a cyclic vector for V. For the case of $I+V$ the proof is similar. $\square$

3. (Non) Supercyclicity.

The adjoint of Volterra operator is defined by

$$V^\star f(x) = \int_x^1 f(s)\,ds,$$

that is, it is an integral operator. It easy to compute that $\sigma_p(V^\star) = \emptyset$. Observe that Volterra operator is defined on complex valued functions. The following result which appear in [**LM**] will reduce our problem to real functions.

Theorem 3.1 (Positive-Supercyclicity's Theorem)**.** *Let T be a bounded linear operator defined on a separable Banach space $\mathcal{B}$. If $\sigma_p(T^\star) = \emptyset$ then T is supercyclic if and only if there exists a vector $x \in \mathcal{B}$ such that $\{rT^n x : r > 0\,, n \in \mathbb{N}\}$ is dense in $\mathcal{B}$.*

Theorem 3.2. *Volterra and the identity plus Volterra operators are not supercyclic on $L^2[0,1]$.*

Proof. Let us denote by $T = V$ or $I + V$. The proof will be done in several steps:

(1) If T is supercyclic on $L^2[0,1]$ then T is supercyclic on $L^2_{\mathbb{R}}[0,1]$.

Proof. Let us denote by $f = u + iv$ a supercyclic vector for T. Observe that $T(L^2_{\mathbb{R}}[0,1]) \subset L^2_{\mathbb{R}}[0,1]$ and $T^n f = T^n u + iT^n v$. It is easy to see (using the positive-supercyclicity's Theorem) that the function u is supercyclic for T on $L^2_{\mathbb{R}}[0,1]$.

(2) If $f \in L^2_{\mathbb{R}}[0,1]$ is a continuous function (more precisely, there exists a continuous function in the coset determined by f) and f is a supercyclic vector for T then the point 0 is an accumulation point of zeros of f.

Proof. Observe that if f is a continuous function so that f is positive (respectively negative) on $[0,\delta]$ then the function $Vf(x)$ is also positive (respectively negative) on $[0,\delta]$. Since Tf is a continuous function we obtain that the orbit under T of f is positive (negative) a.e. $[0,\delta]$. By way of contradiction suppose that $\delta \in (0,1]$ is the smaller zero of f and without loss of generality suppose that f is positive on $(0,\delta)$. In this situation the function -1 is separated more than δ from the set

$$\{cT^n f \;:\; c > 0\,, n \in \mathbb{N}\}.$$

Therefore f cannot be supercyclic for T.

(3) If $f \in L^2_{\mathbb{R}}[0,1]$ is a continuous function, and f is a supercyclic vector for $T^\star$ then the point 1 is an accumulation point of zeros of f.

Proof. The proof of (3) is analogous. It is sufficient to observe that if f is a continuous function on $[0,1]$ and f is positive on $[\delta, 1]$ with $\delta \in [0,1)$ then the orbit under $T^\star$ of f is positive a.e. $[\delta, 1]$.

(4) The operator T is supercyclic if and only if $T^\star$ is supercyclic.

Proof. Let us consider the isomorphism $R : L^2[0,1] \to L^2[0,1]$ defined by $Rf(x) = f(1-x)$. Observe that $T = RT^\star R^{-1}$. Since Supercyclicity is invariant under similarity we obtain (4).

(5) Suppose that V is supercyclic. Then there exists a supercyclic vector f for V which is so that the point 1 is an accumulation point of zeros of $V^n f$ for each integer n. Analagously, if $I+V$ is supercyclic then there exists a supercyclic vector f for $(I+V)$ such that the point 1 is an accumulation point of zeros of the function $V(I+V)^n f$ for each integer n.

Proof. Let us suppose that V is supercyclic, let us denote by G the set of supercyclic vectors for V. It is well-known that the set of supercyclic vectors for a supercyclic bounded linear operator is a G-δ dense subset. By (4) let us denote by $G_\star$ the set of supercyclic vectors for $V^\star$. Since V is continuous the set $V^{-n}(G_\star)$ is also a G-δ dense subset. Therefore the intersection $H = \bigcap_{n=1}^{\infty} V^{-n}(G_\star) \bigcap G$ contains a dense subset. Pick $f \in H$. Clearly f is supercyclic for V, on the other hand if $n \geq 1$, $V^n f \in G_\star$ and $V^n f$ is a continuous function. Therefore by (3) the point 1 is an accumulation point of zeros of $V^n f$.

For the second part let us consider the set $\bigcap_{n=1}^{\infty}(I+V)^{-n}V^{-1}G_\star \bigcap G$ where G and $G_\star$ denote now the sets of supercyclic vectors for $(I+V)$ and $(I+V)^\star$ respectively. The rest of the proof runs as before.

(6) The Volterra and the identity plus Volterra operators are not supercyclic on $L^2_{\mathbb{R}}[0,1]$.

Proof. We first prove that Volterra operator is not supercyclic. It is sufficient to show that the orbit $V^n f$ of a possible supercyclic vector f is orthogonal to the constants, that is, $\langle V^n f, 1\rangle = 0$ for all n. Fix $\epsilon > 0$. If V is supercyclic let us consider the supercyclic function f which guarantee (5). For $n \geq 1$ let us denote by c_n a zero of $V^{n+1} f$ with $c_n \geq 1 - \epsilon$. Since $V^{n+1} f$ is a primitive function of $V^n f$ by applying Barrow's formula we have:

$$\begin{aligned}
|\langle V^n f, 1\rangle|^2 &= \left(\left| \int_0^{c_n} V^n f(s)\, ds \right| + \left| \int_{c_n}^1 V^n f(s)\, ds \right| \right)^2 \\
&= \left| \int_{c_n}^1 V^n f(s)\, ds \right|^2 \\
&\leq (1 - c_n) \int_{c_n}^1 |V^n f(s)|^2 \; ds \\
&\leq (1 - c_n) \|V^n f\|^2 \leq \varepsilon \|V^n f\|^2.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrarily small (and independent of n) we obtain $\langle V^n f, 1\rangle = 0$ for all n, that is f is not cyclic, a contradiction. For the case of $I + V$ the proof is similar.

Thus, by (1) and (6) the proof of Theorem 3.2 is established. $\square$

Observe that although the results are stated in the space $L^2[0,1]$ the proofs runs as well for the spaces $L^p[0,1]$, $1 \leq p < \infty$.

When this paper was being accepted for publication, we were kindly informed by Prof. Joel H. Shapiro about the reference [**GM**] where the authors have obtained the Theorem 3.2 independently.

The authors are deeply grateful to professor Joel Shapiro since the first day that read ours humble work. We want to also thank M. Cepedello Boiso for interesting comments.

References

[Br] M.S. Brodskii, *On a problem of I.M. Gelfand*, Uspekhi Mat. Nak., **12** (1957), 129-132, MR 20 #1229.

[Co] J.B. Conway, *A Course in Operator Theory*, Rhode Island, Amer. Math. Soc., 2000, MR 2001d:47001, Zbl 0936.47001.

[Dix] J. Dixmier, *Les operateurs permutables á l'operateur integral*, Portugal Math., **8** (1949), 73-84, Zbl 0036.35703.

[Don] W.F. Donoghue, *The latice of invariant subspaces of quasi-nilpotent completely continuous transformation*, Pacific J. Math., **7** (1957), 1031-1035, MR 19,1066f, Zbl 0078.29504.

[GM] E. Gallardo and A. Montes, *Volterra operator is not supercyclic*, Integral Equations and Operator Theory, to appear.

[Ha] P.R. Halmos, *A Hilbert Space Problem Book*, New York, Springer Verlag, 1967, MR 34 #8178, Zbl 0144.38704.

[HK] D.A. Herrero and C. Kitai, *On invertible hypercyclic operators*, Proc. Amer. Math. Soc., **116** (1992), 873-875, MR 93a:47023, Zbl 0780.47006.

[HW] H.M. Hilden and L.J. Wallen, *Some cyclic and non-cyclic vectors of certain operators*, Indiana Univ. Math. J., **23** (1974), 557-565, MR 484796, Zbl 0274.47004.

[Ka] G.K. Kalisch, *On similarity, reducing manifolds, and unitary equivalence of certain Volterra operators*, Ann. Math., **66** (1957), 481-494, MR 19,970a, Zbl 0078.09602.

[LM] F. León-Saavedra and V. Müller, *Rotations of hypercyclic and supercyclic operators*, preprint.

[LR] G. Little and J.B. Reade, *Estimates for the norm of the n-th indefinite integral*, Bull. London Math. Soc., **30** (1998), 539-542, MR 99g:47068, Zbl 0931.47032.

[Sa] N.H. Salas, *Supercyclicity and weighted shifts*, Studia Math., **135** (1999), 55-74, MR 2000b:47020, Zbl 0940.47005.

[Sar] D. Sarason, *A remark on the Volterra operator*, J. Math. Anal. Appl., **12** (1965), 244-246, MR 33 #580, Zbl 0138.38801.

[Ts] D. Tsedenbayar, *On the power boundedness of certain Volterra operator pencils*, Studia Math., **156** (2003), 59-66, CMP 1 961 061.

Received August 12, 2002. This work was supported partially by the Grants FQM- 257 and "Vicerrectorado de investigación de U.C.A".

Departamento de Matemáticas
Universidad de Cádiz
C/ Sacramento 82, 11003–Cádiz
Spain
E-mail address: fernando.leon@uca.es

Departamento de Matemáticas
Universidad de Cádiz
C/ Sacramento 82, 11003–Cádiz
Spain
E-mail address: antonio.piqueras@uca.es

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

CONDUCTORS OF WILDLY RAMIFIED COVERS, III

Rachel J. Pries

Consider a wildly ramified G-Galois cover of curves $\phi : Y \to X$ branched at only one point over an algebraically closed field k of characteristic p. In this paper, given G such that the Sylow p-subgroups of G have order p, I show it is possible to deform ϕ to increase the conductor at a wild ramification point. As a result, I prove that all sufficiently large conductors occur for covers $\phi : Y \to \mathbb{P}^1_k$ branched at only one point with inertia $\mathbb{Z}/p$. For the proof, I show there exists such a cover with small conductor under an additional hypothesis on G and then use deformation and formal patching to transform this cover.

1. Introduction.

1.1. Results. Let X be a smooth connected proper curve with marked points $\{x_i\}$ over an algebraically closed field k of characteristic p. Consider a Galois cover $\phi : Y \to X$ of smooth connected curves branched only at $\{x_i\}$. Abhyankar's Conjecture (proved by Raynaud [**10**] and Harbater [**4**]) determines exactly which groups G can be the Galois group of ϕ. An open problem is to determine which inertia groups and filtrations of higher ramification groups can be realized for such a cover ϕ. More simply, it is unknown which integers can be realized as the genus of Y.

The main results of this paper are for the case that $\phi : Y \to \mathbb{P}^1_k$ is branched at only one point. Such covers exist if and only if G is a quasi-p group, which means that G is generated by p-groups. Harbater [**3**] proved that the Sylow p-subgroups of G can be realized as the inertia groups of such a cover ϕ. Under the assumption that the Sylow p-subgroups of G have order p, the filtration of higher ramification groups is determined by one integer j for which $p \nmid j$, namely by the lower jump or conductor. In Theorem 3.2.4, I prove in this case that all sufficiently large conductors occur for such covers of the affine line. Theorem 3.2.4 involves the concept of the *p-weight* which is defined in Section 3.1.

Theorem 3.2.4. *Let G be a finite quasi-p group whose Sylow p-subgroups have order p. There exists an integer J depending explicitly on p, the p-weight of G, and the exponent of the normalizer of a Sylow p-subgroup of G with the following property: If $j \geq J$ and $p \nmid j$ then there exists a G-Galois* *cover $\phi : Y \to \mathbb{P}^1_k$ branched at only one point over which it has inertia group $\mathbb{Z}/p$ and conductor j.*

The first part of the proof is to show that all sufficiently large conductors will occur. To do this, I show the following more general result in Theorem 2.2.2: Suppose $\phi : Y \to X$ is a G-Galois cover with inertia group I of the form $\mathbb{Z}/p \rtimes \mu_m$ and conductor j at a ramification point; then it is possible to deform ϕ to increase this conductor. To do this, I construct a family of covers so that ϕ is isomorphic to the normalization of one fibre of the family. The techniques consist of local deformations and formal patching theorems of Harbater and Stevenson [**5**]. Since k is algebraically closed, it is then possible to use another fibre of this family to find another cover with the same group G and inertia group I but with a larger conductor. For certain applications, it is necessary to use ramification data of two covers and deform semi-stable curves in order to enlarge the Galois group and to change the inertia group, while simultaneously enlarging the conductor; see Theorem 2.3.7.

The second part of the proof is to find a relatively small integer J (depending only on the group theory of G) for which there exists a G-Galois cover $\phi : Y \to \mathbb{P}^1_k$ branched at only one point over which it has inertia group $\mathbb{Z}/p$ and conductor J. For this I use the following result which says roughly speaking that there exists such a cover of the affine line with very small conductor when G has p-weight one. (See 3.2.1 for the definition of $j_{\min}(I)$, which is a small set of integers depending only on I and not on G consisting of the minimal possible conductors for a cover of the affine line with inertia $I \simeq \mathbb{Z}/p \rtimes \mu_m$.)

Theorem 3.2.2. *Let G be a finite quasi-p group of p-weight one whose Sylow p-subgroups have order $p \neq 2$. For some $I \simeq \mathbb{Z}/p \rtimes \mu_m \subset G$ and some $j \in j_{\min}(I)$, there exists a G-Galois cover $\phi : Y \to \mathbb{P}^1_k$ of smooth connected curves branched at only one point over which it has inertia group I and conductor j. In particular,* $\text{genus}(Y) \leq 1 + |G|(p-1)/2p$.

This result was announced in [**7**]. The idea behind its proof is to reverse the process in Section 2 to decrease the conductor of a G-Galois cover $\phi : Y \to \mathbb{P}^1_k$ branched at only one point. This is done by analyzing the stable model of a family of covers with bad reduction over an equal characteristic discrete valuation ring. This is motivated by the work of Raynaud [**11**] in unequal characteristic.

1.2. Notation and background. Let k be an algebraically closed field of characteristic p. Let $R \simeq k[[t]]$ be an equal characteristic complete discrete valuation ring with residue field k and fraction field $K \simeq k((t))$. For each $m \in \mathbb{N}$ with $\gcd(m, p) = 1$, choose an mth root of unity $\zeta_m \in k$ such that

$\zeta_{m_2} = \zeta_{m_2 m_1}^{m_1}$. Let G be a finite group and let S be a chosen Sylow p-subgroup of G. The group G is *quasi-p* if it is generated by all its Sylow p-subgroups.

If X is a scheme over R, we assume that the morphism $f : X \to \mathrm{Spec}(R)$ is separated, flat and of finite type. If ξ is a point of a scheme X, the *germ* $\hat{X}_\xi$ of X at ξ is defined to be the spectrum of the complete local ring of functions of X at ξ. Suppose a scheme X is reduced and connected, but not necessarily irreducible. A morphism $\phi : Y \to X$ of schemes is a (possibly branched) *cover* if ϕ is finite and generically separable. A *G-Galois cover* is a cover $\phi : Y \to X$ along with a choice of homomorphism $G \to \mathrm{Aut}_X(Y)$ by which G acts simply transitively on each generic geometric fibre of ϕ (again allowing branching). If $\phi : Y \to X$ is a G-Galois cover and $G \subset G'$, define the *induced cover* $\mathrm{Ind}_G^{G'}(Y) \to X$ to be the disconnected G'-Galois cover consisting of $(G' : G)$ copies of Y indexed by the left cosets of G with the induced action of G'.

Consider a wildly ramified G-Galois cover of curves $\phi : Y \to X$ with branch locus B. See [**13**, Chapter IV] for information about the higher ramification groups of ϕ. In particular, suppose $\xi \in B$ is a closed point and $\eta \in \phi^{-1}(\xi)$. The inertia group I of ϕ at η is of the form $I = P \rtimes \mu_m$ where P is a p-group and $p \nmid m$. In the case that $I \simeq \mathbb{Z}/p \rtimes \mu_m$, the *conductor* of ϕ at η is the integer $j = \mathrm{val}(q(\pi_\eta) - \pi_\eta) - 1$ where π_η is a uniformizer of Y at η and q has order p in I. In this case, the conductor j is the unique *lower jump* in the filtration of higher ramification groups and the *upper jump* is $\sigma = j/m \in \mathbb{Q}$. Up to isomorphism, these objects do not depend on the choice of η above ξ. If ξ is not a closed point of X, the inertia group, filtration of higher ramification groups, and conductor for ϕ at η are the corresponding objects over the generic point of η.

This paper frequently uses the following technique of Harbater and Stevenson [**5**] (also see [**9**] for the case that $R \not\simeq k[[t]]$). Let X be a projective k-curve that is connected and reduced but not necessarily irreducible. Let $\mathbb{S}$ be a finite closed subset of X which contains the singular locus of X.

Definition 1.2.1. A *thickening problem of covers* for $(X, \mathbb{S})$ consists of the following data:

(a) A cover $f : Y \to X$ of geometrically connected reduced projective k-curves;
(b) for each $s \in \mathbb{S}$, a Noetherian normal complete local domain R_s containing R such that t is contained in the maximal ideal of R_s, along with a finite generically separable R_s-algebra A_s;
(c) for each $s \in \mathbb{S}$, a pair of k-algebra isomorphisms $F_s : R_s/(t) \to \hat{\mathcal{O}}_{X,s}$ and $E_s : A_s/(t) \to \hat{\mathcal{O}}_{Y,s}$ which are compatible with the inclusion morphisms.

Definition 1.2.2. A thickening problem is *G-Galois* if f and $R_s \subset A_s$ are G-Galois and the isomorphisms F_s are compatible with the G-Galois action (for all $s \in \mathbb{S}$). A *thickening* of X is a projective normal R-curve X^* such that $X_k^* \simeq X$. A thickening problem is *relative* if the data for the problem also includes a thickening X^* of X, so that X^* is a trivial deformation of X away from $\mathbb{S}$ and so that the pullback of X^* over the complete local ring at a point $s \in \mathbb{S}$ is isomorphic to R_s.

Definition 1.2.3. A *solution* to a thickening problem of covers is a cover $f^* : Y^* \to X^*$ of projective normal R-curves, where X^* is a thickening of X, whose closed fibre is isomorphic to f, whose pullback to the formal completion of X^* along $X' = X - \mathbb{S}$ is a trivial deformation of the restriction of f over X', and whose pullback over the complete local ring at a point $s \in \mathbb{S}$ is isomorphic to $R_s \subset A_s$ (and such that everything is compatible with the isomorphisms above). (Note that X^* is a thickening of X.)

Theorem 1.2.4 (Harbater, Stevenson)**.** *Every (G-Galois) thickening problem for covers has a (G-Galois) solution. The solution is unique if the thickening problem is relative.*

Proof. [**5**, Theorem 4]. □

2. Deformation of covers.

Consider a G-Galois cover $\phi : Y \to X$ of smooth connected proper k-curves. Let ξ be in the branch locus of ϕ and let $\eta \in \phi^{-1}(\xi)$. The goal in this section is to deform the cover ϕ with precise control over the ramification behavior near ξ. To do this it is first necessary to deform the I-Galois cover $\hat{\phi} : \hat{Y}_\eta \to \hat{X}_\xi$ of germs of curves near ξ with such control. We assume throughout that p strictly divides the order of the inertia group I.

2.1. Covers of complete local rings. Let I be a semi-direct product $\mathbb{Z}/p \rtimes \mu_m$ with $p \nmid m$. Let n' be the order of the prime-to-p part of the center of I. Let $U = \mathrm{Spec}(k[[u]])$ and let b be the closed point of U. The next results describe the structure of I-Galois covers $\phi : X \to U$ of germs of curves with lower jump j in the filtration of higher ramification groups above b.

Definition 2.1.1. Suppose $\phi : X \to U$ is an I-Galois cover of germs of curves such that X is connected but not necessarily normal. Let r_1 be the number of connected components of the normalization of X and assume that $p \nmid r_1$. Define the *inertia group* of ϕ to be the inertia group I_1 of a closed point in the normalization. The order of I is $pm = pm_1r_1$ where pm_1 is the order of I_1. The *conductor* (respectively *upper jump*) of ϕ is defined to be the conductor (respectively upper jump) of a ramification point in the normalization.

Lemma 2.1.2. *Let $I \simeq \mathbb{Z}/p \rtimes \mu_m$ and let $I_1 \subset I$ have index r_1 with $p \nmid r_1$. Write $m = r_1 m_1$. Suppose that $\phi : X \to U$ is an I-Galois cover of connected germs of curves with inertia I_1 and conductor j.*

i) *There exists an automorphism A of U such that the equations for $A^*\phi$ away from b are:*

$$u_1^m = u^{r_1},\ x^p - x = u_1^{-j}.$$

ii) *The Galois action on the generic fibre is given by the following equations for some γ with* $\gcd(\gamma, m) = 1$; (*after possibly changing the choice of q*)*:*

$$c(u_1) = \zeta_m^{\gamma} u_1,\ c(x) = \zeta_m^{-\gamma j} x,\ q(u_1) = u_1,\ q(x) = x + 1.$$

iii) *The conductor j satisfies $p \nmid j$ and* $\gcd(j, m) = n'$. *The upper jump is $\sigma = j/m_1$.*

Proof. First consider the case that Y is normal (i.e., $r_1 = 1$). By [**8**, Lemma 1.4.2], there exists $f(u) \in k[[u]]^*$ with degree j so that the equations for ϕ away from $u = 0$ are $u_1^m = u$ and $x^p - x = u_1^{-j} f(u)$. The proof uses Kummer theory and Artin-Schreier Theory. To finish the proof of i) consider the automorphism A of $k[[u]]$ such that $A(u) = u f(u)^{m/j}$. This automorphism exists since $f(u)$ is a jth power in $k[[u]]^*$. Then $A(u_1) = u_1 f(u)^{1/j}$. After this automorphism of the base, the equations for $A^*\phi$ are $u_1^m = u$ and $x^p - x = u_1^{-j}$.

Now consider the case that $r_1 \neq 1$. The normalization of ϕ is a disconnected cover whose components are Galois with group I_1. Thus the normalization has equations:

$$v^{r_1} = 1,\ u_1^{m_1} = vu,\ x^p - x = u_1^{-j}.$$

The equations for $A^*\phi$ in i) are a blow-down of these by the relation $uv = u_1^{m_1}$. Properties ii)-iii) follow directly from [**13**, Chapter IV] and [**8**, Lemma 1.4.2]. □

The next lemma allows one to induce a given I_1-Galois cover up to a reducible connected I-Galois cover if the restriction from Lemma 2.1.2 (iii) is satisfied. This will be used in Proposition 2.3.4 to patch together covers with different inertia groups.

Lemma 2.1.3. *Suppose $I_1 \subset I \simeq \mathbb{Z}/p \rtimes \mu_m$ with index r_1 where $p \nmid r_1$. Let $m = m_1 r_1$. Suppose there exists an I_1-Galois cover ϕ of connected normal germs of curves with conductor j. Assume $n' =$ $\gcd(m, j)$. Then there exists a connected reducible I-Galois cover ϕ^{ind} with conductor j which is isomorphic to* $\mathrm{Ind}_{I_1}^{I}(\phi)$ *away from the closed point.*

Proof. Let $c_1 = c^{r_1}$. By Lemma 2.1.2, there is an automorphism A of U so that the equations for $A^*\phi$ are $u_1^{m_1} = u$, $x^p - x = u_1^{-j}$; and its I_1-Galois action is given by $c_1(u_1) = \zeta_{m_1}^{\gamma} u_1$, $c_1(x) = \zeta_{m_1}^{-\gamma j} x$, $q(u_1) = u_1$ and $q(x) = x+1$ for some γ with $\gcd(\gamma, m_1) = 1$. The equations for $\mathrm{Ind}_{I_1}^{I}(A^*\phi)$ are $v^{r_1} = 1$, $u_1^{m_1} = uv$ and $x^p - x = u_1^{-j}$.

Let ϕ_A^{ind} be the blow down of $\mathrm{Ind}_{I_1}^{I}(A^*\phi)$ which identifies the r_1 ramification points. This yields a connected reducible I-Galois cover ϕ_A^{ind} whose equations and I-Galois action are the same as in Lemma 2.1.2 (i) and (ii). This Galois action is well-defined by the condition on n'. Let $\phi^{\mathrm{ind}} = (A^{-1})^*\phi_A^{\mathrm{ind}}$ and note that ϕ^{ind} is isomorphic to $\mathrm{Ind}_{I_1}^{I}(\phi)$ away from the closed point by construction. □

2.2. Deformation of smooth curves. In this section, we show that it is possible to increase the conductor at a branch point while preserving the inertia and Galois group. This result was announced in [**7**]. Let $R = k[[t]]$ and $K = k((t))$. Let b be the closed point of $U = \mathrm{Spec}(k[[u]])$. Let $U_R = \mathrm{Spec}(R[[u]])$ and $U_K = U_R \times_R K = \mathrm{Spec}(k[[u,t]][t^{-1}])$.

Proposition 2.2.1. *Let $I \simeq \mathbb{Z}/p \rtimes \mu_m$. Suppose there exists an I-Galois cover $\phi : X \to U$ of normal connected germs of curves with conductor j. Then for $i \in \mathbb{N}$ with $p \nmid (j + im)$, there exists an I-Galois cover $\phi_R : X_R \to U_R$ of irreducible germs of R-curves, whose branch locus consists of only the R-point $b_R = b \times_k R$, such that:*

1. *The normalization of the special fibre of ϕ_R $(t = 0)$ is isomorphic to ϕ away from b.*
2. *The generic fibre $\phi_K : X_K \to U_K$ of ϕ_R is an I-Galois cover of normal connected curves whose branch locus consists of only the K-point $b_K = b_R \times_R K$ over which it has inertia I and conductor $j + im$.*

Proof. After an automorphism A of $k[[u]]$, the equations for $A^*\phi$ are given by: $u_1^m = u$, $x^p - x = u_1^{-j}$. Consider the normal cover $\phi'_R : X'_R \to U_R$ given generically by the equations:

$$u_1^m = u, x^p - x = u_1^{-(j+im)}(t + u^i).$$

The I-Galois action on the variables is given by the same expressions as on the closed fibre and the cover is irreducible. The curve X'_R is singular only above the point $(u,t) = (0,0)$. The normalization of the special fibre agrees with $A^*\phi$. The cover ϕ'_R is branched only at the R-point $u = 0$ since $u_1 = 0$ is the only pole of the function $u_1^{-(j+im)}(t + u^i)$. Taking the restriction of ϕ_K over $\mathrm{Spec}(K[[u]])$ where $t + u^i$ is a unit, we see that ϕ_K has inertia I and conductor $j + im$ over b_K. Pulling back the cover ϕ'_R by the automorphism $A^{-1} \times_k R$ of $R[[u]]$ changes none of these properties and thus yields the cover ϕ_R. □

Let X_k be a proper k-curve. The next theorem uses Propositon 2.2.1 and Theorem 1.2.4 to deform a given cover of X_k to a family of covers ϕ_R of X_k. This family can be defined over a variety Θ of finite type over k. We then specialize to a fibre of the family over another k-point of Θ to get a cover ϕ' with new ramification data.

Theorem 2.2.2. *Suppose there exists a G-Galois cover $\phi : Y \to X_k$ of smooth connected curves with branch locus B. Suppose ϕ has inertia group $I \simeq \mathbb{Z}/p \rtimes \mu_m$ and conductor j above $\xi_1 \in B$ with $p \nmid m$. Let $i \in \mathbb{N}^+$ be such that $p \nmid (j+im)$. Then there exist G-Galois covers $\phi_R : Y_R \to X_R$ and $\phi' : Y' \to X_k$ such that:*

1. *The curves Y_R and Y' are irreducible and Y_K and Y' are smooth and connected.*
2. *After normalization, the special fibre ϕ_k of ϕ_R is isomorphic to ϕ away from ξ_1.*
3. *The branch locus of the cover ϕ_R (respectively ϕ') consists exactly of the R-points $\xi_R = \xi \times_k R$ (respectively the k-points ξ) for $\xi \in B$.*
4. *For $\xi \in B$, $\xi \neq \xi_1$, the ramification behavior for ϕ_R (respectively ϕ') at ξ_R (respectively ξ) is identical to that of ϕ at ξ.*
5. *The cover ϕ_K (respectively ϕ') has inertia I and conductor $j+im$ at the K-point $\xi_{1,K}$ (respectively at ξ_1).*
6. *The genus of Y' and of the fibres of Y_R is $g'_Y = \mathrm{genus}(Y) + i|G|(p-1)/2p$.*

Proof. In the notation of Theorem 1.2.4, let $X^* = X_R$ and let $\mathbb{S} = \{\xi_1\}$. Let $\eta \in \phi^{-1}(\xi_1)$. Consider the I-Galois cover $\hat{\phi} : \hat{Y}_\eta \to \hat{X}_{\xi_1}$. Applying Proposition 2.2.1 to $\hat{\phi}$, there exists a deformation $\hat{\phi}_R : \hat{Y}_R \to \hat{X}_R$ of $\hat{\phi}$ with the desired properties. In particular, $\hat{\phi}_K$ has inertia $I \simeq \mathbb{Z}/p \rtimes \mu_m$ and conductor $j+im$ over $\xi_{1,K}$. Consider the inclusion $R_s \to A_s$ of rings corresponding to the disconnected G-Galois cover $\mathrm{Ind}_I^G(\hat{\phi}_R)$.

The covers ϕ_k and $\mathrm{Ind}_I^G(\hat{\phi}_R)$ and the isomorphism given by Proposition 2.2.1 (1) constitute a relative G-Galois thickening problem as in Definition 1.2.1. The (unique) solution to this thickening problem (Theorem 1.2.4) yields the G-Galois cover $\phi_R : Y_R \to X_R$. Recall from Definition 1.2.3 that the cover ϕ_R is isomorphic to $\mathrm{Ind}_I^G(\hat{\phi}_R)$ over $\hat{X}_{R,\xi_1}$. Also, ϕ_R is isomorphic to the trivial deformation $\phi_{tr} : Y_{tr} \to X_{tr}$ of ϕ away from ξ_1. Thus Y_R is irreducible since Y is irreducible and Y_K is smooth since $Y_{tr,K}$ and $\hat{Y}_K$ are smooth.

The data for the cover ϕ_R is contained in a subring $\Theta \subset R$ of finite type over k, with $\Theta \neq k$ since the family is nonconstant. Since k is algebraically closed, there exist infinitely many k-points of $\mathrm{Spec}(\Theta)$. The closure L of the locus of k-points x of $\mathrm{Spec}(\Theta)$ over which the fibre ϕ_x is not a G-Galois cover of smooth connected curves is closed, [**2**, Proposition 9.29]. Furthermore,

$L \neq \mathrm{Spec}(\Theta)$ since Y_K is smooth and irreducible. Let $\phi' : Y' \to X_k$ be the fibre over a k-point not in L. Note that Y' is smooth and irreducible by definition. The other properties follow immediately from the compatibility of ϕ_R with $\mathrm{Ind}_I^G(\hat{\phi}_R)$ over $\hat{X}_{R,\xi_1}$ and with $\phi_{tr} : Y_{tr} \to X_{tr}$ away from ξ_1.

The genus of Y' and of the fibres of Y_R increases because of the extra contribution to the Riemann-Hurwitz formula. In particular, there are $|G|/mp$ ramification points above $\xi_{1,K}$, each of which has im extra nontrivial higher ramification groups. Thus the degree of ramification $\mathrm{Deg}(\xi_1)$ over $\xi_{1,K}$ increases by $|G|(im)(p-1)/mp$. □

Theorem 2.2.2 can be used to increase the conductor of a cover of proper curves while preserving the inertia and Galois group. For some applications, however, it is necessary to change the Galois group, the prime-to-p part of the inertia group, or the congruence value of the conductor. To do this, it is necessary to deform covers of semistable curves.

2.3. Deformation of semi-stable curves. In this section, we deform covers of semi-stable curves with control over the ramification information. The motivation for Theorem 2.3.7 is that it allows us to use two wildly ramified covers to produce another whose Galois group, inertia group and conductor at a point are determined from the given ones. In Section 3, we use this theorem to produce a cover with complicated Galois group and relatively small conductor. See [**1**] for an application in which the prime-to-p part of the inertia group, and the congruence value of the conductor are changed using this theorem. Since the notation involved in Theorem 2.3.7 is complicated, we will start with a corollary.

Corollary 2.3.1. *Let G be a group generated by subgroups G_1 and G_2. Let $\phi_1 : X \to \mathbb{P}^1_k$ and $\phi_2 : Y \to \mathbb{P}^1_k$ be Galois covers of smooth connected curves each branched at only one point. Suppose ϕ_i has group G_i, inertia I_i, and conductor j_i respectively. Then there exists a G-Galois cover $\phi' : Y' \to \mathbb{P}^1_k$ of smooth connected curves, branched at only one point with inertia I' and conductor j' as follows:*

1. **Purely wild:** *Suppose $I_1 = I_2 = \mathbb{Z}/p$. Then $I' = \mathbb{Z}/p$ and $j' = j_1 + j_2 + \epsilon$. Here $\epsilon = 0$ if $p \nmid (j_1 + j_2)$; if $p | (j_1 + j_2)$, then $\epsilon = 2$ if $p \neq 2$ and $\epsilon = 3$ if $p = 2$.*
2. **Admissible:** *Suppose $I_1 \simeq I_2$ and $j_1 = \gamma j_2$ for some $\gamma \equiv -1 \bmod m$. Write $\gamma = \nu m - 1$. Assume that $p \nmid (j_1 + j_2)$. Then $I' = \mathbb{Z}/p$ and $j' = (j_1 + j_2)/m = \nu j_2$.*
3. **Different inertia:** *More generally, suppose $I_1 \subset I_2 = \mathbb{Z}/p \rtimes \mu_m$ with index r for some r with $p \nmid r$. Let $e = j_1 r + j_2$ and let $g = \gcd(m, e/\gcd(j_1, j_2))$. Assume that $j_1 = \gamma j_2$ for some γ such that $\gcd(\gamma, m) = 1$. Assume that $p \nmid e$. Then I' is the unique subgroup of I_2 with order pm/g and $j' = e/g$.*

The second case is called admissible since the prime-to-p ramification disappears.

Proof. The proof is immediate from Theorem 2.3.7 and Proposition 2.3.5. $\square$

Notation 2.3.2. Let $U = \mathrm{Spec}(k[[u]])$ and $V = \mathrm{Spec}(k[[v]])$. For $e \in \mathbb{N}^+$, we define $\Omega^e_{uv} = k[[u, v, t]]/(uv - t^e)$ and let $S^e_{uv} = \mathrm{Spec}(\Omega^e_{uv})$. Let $\iota_u : U \to S^e_{uv}$ and let $\iota_v : V \to S^e_{uv}$ be the natural inclusions. Let $b \in S^e_{uv}$ be the k-point $(u, v, t) = (0, 0, 0)$. Suppose $I \simeq \mathbb{Z}/p \rtimes \mu_m = \mathbb{Z}/p \rtimes \langle c\rangle$. Let n' be the order of the prime-to-p part of the center of I. For $i = 1, 2$, suppose $I_i = \mathbb{Z}/p \rtimes \mu_{m_i} \subset I$ with index r_i where $p \nmid r_i$. Note that $m = m_1r_1 = m_2r_2$.

Let $\phi_1 : X \to U$, $\phi_2 : Y \to V$ be Galois covers of normal connected germs of curves. Suppose the cover ϕ_i has inertia (and Galois) group $I_i \simeq \mathbb{Z}/p \rtimes \mu_{m_i}$ and conductor j_i for $i = 1, 2$. Let $g' = \gcd(j_1, j_2)$. Let $e = j_1r_1 + j_2r_2$, and let $g = \gcd(m, e/g')$. Let $e' = e/g'$, $j'_1 = j_1/g'$ and $j'_2 = j_2/g'$.

Numerical hypotheses: Suppose that $n' = \gcd(m, j_1) = \gcd(m, j_2)$. Suppose $p \nmid e$. Suppose that $j'_1 \equiv \gamma j'_2 \bmod m$ for some γ such that $\gcd(\gamma, m) = 1$ and $1 \leq \gamma < m$. Suppose that $\gcd(j'_2, m) = 1$.

The first condition is necessary to dominate each cover by an I-Galois cover. The other three conditions imply that $p \nmid (e/g)$; $\gcd(j'_1, m) = 1$; and $g = \gcd(m, \gamma r_1 + r_2)$. Also, when $j_1 = \gamma j_2$ and $p \nmid (\gamma + 1)$ for some γ with $\gcd(\gamma, m) = 1$ then the three last conditions are satisfied.

Proposition 2.3.4 constructs an I-Galois cover $\phi_R : W_R \to S^{e'}_{uv}$ with specified ramification from the covers ϕ_1 and ϕ_2. Although W_R will be flat over R and normal, its special fibre W_k will be singular at the point $w = \phi_R^{-1}(b)$.

The following lemma will be used in the proof of Proposition 2.3.4:

Lemma 2.3.3. *Suppose $\ell = \ell_1 + \ell_2$ with $\ell_i \in \mathbb{N}^+$ and $a \in k$. Then $d_0 \in \Omega^\ell_{uv}$ where*
$$d_0 = \frac{(u + at^{\ell_2})^\ell - u^\ell - (at^{\ell_2})^\ell}{u^{\ell_2}t}.$$

Proof. It is sufficient to show that the binomial coefficient $c_i = u^it^{\ell_2(\ell-i)}/u^{\ell_2}t \in \Omega^\ell_{uv}$ for $1 \leq i \leq \ell - 1$. Since $t^\ell = uv$,
$$c_i = u^{i-\ell_2}(uv)^{\ell_2}t^{-\ell_2 i-1} = u^iv^{\ell_2}t^{-\ell_2 i-1}.$$
Since Ω^ℓ_{uv} is normal, it is sufficient to check that $c_i^\ell \in \Omega^\ell_{uv}$. Here
$$c_i^\ell = u^{\ell i}v^{\ell\ell_2}(uv)^{-\ell_2 i-1} = u^{i(\ell-\ell_2)-1}v^{\ell_2(\ell-i)-1}.$$
Thus $c_i^\ell \in \Omega^\ell_{uv}$ since $\ell - \ell_2 \geq 1$ and $\ell - i \geq 1$. $\square$

Proposition 2.3.4. *Consider the pair (ϕ_1, ϕ_2) from in Notation* 2.3.2 *satisfying the numerical hypotheses. There exists an I-Galois cover $\phi_R : W_R \to S^{e'}$ of* (*possibly disconnected*) *germs of R-curves and an isomorphism $i : S^{e'}_{uv} \to S^{e'}$ such that:*

1. *The branch locus of ϕ_R consists of one R-point, denoted b_R, which specializes to b.*
2. *After normalization, the pullbacks of the special fibre of ϕ_R to U and V, namely $\iota_u^* i^* \phi_k$ and $\iota_v^* i^* \phi_k$, are isomorphic to a disjoint union of copies of respectively ϕ_1 and ϕ_2 away from the branch point b.*
3. *The generic fibre $\phi_K : X_K \to S^{e'}_K = S^{e'} \times_R K$ of ϕ_R is an I-Galois cover of* (*possibly disconnected*) *germs of curves branched at exactly the K-point $b_K = b_R \times_R K$.*
4. *The cover ϕ_K has g ramification points above b_K, each with inertia group $I_K \simeq \mathbb{Z}/p \rtimes \langle c^g \rangle = \mathbb{Z}/p \rtimes \mu_{m/g}$ and conductor $e/g = (j_1 r_1 + j_2 r_2)/g$.*
5. *The curve W_R is irreducible if and only if* $\gcd(r_1, r_2) = 1$. *The curve W_K is irreducible if and only if $g = 1$.*

Proof. The proof is to construct ϕ_R and then verify its properties.

The equations for ϕ_1^{ind} and ϕ_2^{ind}: Applying Lemma 2.1.2, there exist automorphisms A_u of $k[[u]]$ and A_v of $k[[v]]$ which fix the closed points of U and V and such that the pullbacks $A_u^* \phi_1$ and $A_v^* \phi_2$ are given by the equations in Lemma 2.1.2 (i).

Since $n' = \gcd(m, j_1) = \gcd(m, j_2)$, Lemma 2.1.3 implies that there exist connected reducible I-Galois covers ϕ_1^{ind} and ϕ_2^{ind} which are isomorphic respectively to $A_u^* \phi_1$ and $A_v^* \phi_2$ away from the branch point. The equations for these covers are:

$$\phi_1^{\mathrm{ind}} : u_1^m = u^{r_1},\ x^p - x = u_1^{-j_1};$$
$$\phi_2^{\mathrm{ind}} : v_1^m = v^{r_2},\ y^p - y = v_1^{-j_2}.$$

After possibly changing c and q once and for all, the I-Galois action of ϕ_1^{ind} and ϕ_2^{ind} is given (for some γ such that $\gcd(\gamma, m) = 1$ and some $a \in \mathbb{F}_p$) by:

$$c(u_1) = \zeta_m u_1, \quad c(x) = \zeta_m^{-j_1} x, \quad q(x) = x + 1,$$
$$c(v_1) = \zeta_m^{\gamma} v_1, \quad c(y) = \zeta_m^{-\gamma j_2} y, \quad q(y) = y + a.$$

Note that the normalization of ϕ_1^{ind} (resp. ϕ_2^{ind}) is isomorphic to a disjoint union of copies of $A_u^* \phi_1$ (resp. $A_v^* \phi_2$) away from the closed point.

The equations for ϕ_R: Let $g' = \gcd(j_1, j_2)$ and $e' = j_1' r_1 + j_2' r_2$ be as in Notation 2.3.2. There exists $a_1 \in k$ such that $a_1^{g'} = a^m$ since k is algebraically closed. Consider the cover $\phi'_R : W_R \to S^{e'}_{uv}$ given by:

$$w_1^m = u^{j'_1 r_1} + a_1 v^{j'_2 r_2} + d_0 t,\ z^p - z = (1 + d_1 t) w_1^{-g'}.$$

For any choice of the variables $d_0, d_1 \in \Omega^{e'}_{uv}$, the cover ϕ'_R reduces to ϕ_1^{ind} and ϕ_2^{ind} on the components of the special fibre. To see this, note that $\mathrm{mod}(v,t)$ the equations for $\iota_u^* i^* \phi'_R$ are $w_1^m = u^{j'_1 r_1}$ and $z^p - z = w_1^{-g'}$. After making the identification $w_1 \mapsto u_1^{j'_1}$ and $z \mapsto x$, a normalization of these equations is isomorphic to ϕ_1^{ind}. (Specifically, we take a normalization of $u_1^{j'_1 m} = u^{j'_1 r_1}$ and $x^p - x = u_1^{-g' j'_1} = u_1^{-j_1}$.) Likewise, $\mathrm{mod}(u,t)$ the equations for $\iota_v^* i^* \phi'_R$ are $w_1^m = a_1 v^{j'_2 r_2}$ and $z^p - z = w_1^{-g'}$. After making the identification $w_1 \mapsto a_1^{1/m} v_1^{j'_2}$ and $z \mapsto y/a$, this simplifies to $v_1^{j'_2 m} = v^{j'_2 r_2}$ and $y^p - y = a(a_1^{1/m} v_1^{j'_2})^{-g'} = v_1^{-j_2}$. A normalization of these equations is isomorphic to ϕ_2^{ind}. By the numerical hypotheses that $j'_1 \equiv \gamma j'_2 \bmod m$ and $\gcd(j'_1, m) = 1$, there is a well-defined I-Galois action on ϕ'_R which reduces correctly:

$$c(w_1) = \zeta_m^{j'_1} w_1,\ c(z) = \zeta_m^{-j_1} z,\ q(z) = z + 1.$$

In conclusion, after normalization, the pullbacks of the special fibre of ϕ'_R to U and V, namely $\iota_u^* i^* \phi_k$ and $\iota_v^* i^* \phi_k$, are isomorphic to a disjoint union of copies of respectively $A_u^* \phi_1$ and $A_v^* \phi_2$ away from the branch point b.

The cover ϕ_R will be the composition of ϕ'_R with a change of base. Namely, suppose $A_u^{-1}(u) = u d_u$ and $A_v^{-1}(v) = v d_v$ for $d_u \in k[[u]]^*$ and $d_v \in k[[v]]^*$. Recall that $\Omega^{e'}_{uv} = k[[u,v,t]]/(uv - t^{e'})$. Let $\Omega^{e'} = k[[u,v,t]]/(uv d_u d_v - t^{e'})$ and let $S^{e'} = \mathrm{Spec}(\Omega^{e'})$. There exists an isomorphism $A : \Omega^{e'} \to \Omega^{e'}_{uv}$ which reduces to A_u on U and A_v on V.

Consider the pullback of the cover ϕ'_R by A. In other words, consider the cover $\phi_R : W_R \to S^{e'}$ corresponding to the composition $\Omega^{e'} \xrightarrow{A} \Omega^{e'}_{uv} \to \mathcal{O}_{W'}$. Most properties for ϕ_R in Proposition 2.3.4 are automatic by the construction of ϕ'_R. Since A is an isomorphism, to finish the proof it is sufficient to verify Properties 3)-5) for ϕ'_R.

The branch locus: Note that $e' = j'_1 r_1 + j'_2 r_2$ is the sum of two positive integers. Let

$$d_0 = \frac{(u + a_2 t^{j'_2 r_2})^{e'} - u^{e'} - a_1 t^{j'_2 r_2 e'}}{u^{j'_2 r_2} t}.$$

By Lemma 2.3.3, $d_0 \in \Omega^{e'}_{uv}$. Rewrite the first equation for ϕ'_R as:

$$w_1^m = u^{-j'_2 r_2}(u^{e'} + a_1 (uv)^{j'_2 r_2} + d_0 t u^{j'_2 r_2}).$$

Since $uv = t^{e'}$, this simplifies to: $w_1^m = u^{-j'_2 r_2}(u^{e'} + a_1 (t^{j'_2 r_2})^{e'} + d_0 t u^{j'_2 r_2})$.

For this choice of d_0 and for $a_2 = a_1^{1/e}$, the equations on the generic fibre are:

$$w_1^m = u^{-j_2' r_2}(u + a_2 t^{j_2' r_2})^{e'}, \; z^p - z = (1 + d_1 t) w_1^{-g'}.$$

Note that in $\Omega^{e'}_{uv,K}$ the function u has no zero or pole and $1 + d_1 t$ has no poles. Thus ϕ_K' has a unique branch point given by the coordinates $u = -a_2 t^{j_2' r_2}$ and $v = -(t^{e' - j_2' r_2})/a_2 = -t^{j_1' r_1}/a_2$. This K-point specializes to the branch point $(u, v) = (0, 0)$ of ϕ_k. Thus ϕ_R is branched at exactly one R-point for this choice of d_0.

Irreducibility: Consider the equations for the cover ϕ_R':

$$w_1^m = u^{-j_2' r_2}(u + a_2 t^{j_2' r_2})^{e'}, \; z^p - z = (1 + d_1 t) w_1^{-g'}.$$

Note that if m, e' and $j_2' r_2$ share no common factors then the first equation is irreducible; the second is irreducible since the right-hand side is not of the form $\alpha^p - \alpha$. Recall that j_1' and j_2' are relatively prime to m and thus to r_2. Since $\gcd(m, j_2' r_2) = r_2$, the curve W_R is irreducible if and only if $1 = \gcd(e', r_2) = \gcd(r_1, r_2)$. Let $g = \gcd(m, j_1' r_1 + j_2' r_2) = \gcd(m, e')$. We see that W_K is irreducible if and only if $g = 1$ from the following equations for the normalization of ϕ_K':

$$w_2^g = 1, \; w_1^{m/g} = u^{-j_2' r_2/g}(u + a_2 t^{j_2' r_2})^{e'/g}, \; z^p - z = (1 + d_1 t) w_1^{-g'}.$$

Ramification information: The first equation indicates that normalization of the generic fibre has g components and thus g points above the branch point each with inertia group $I_K = \mathbb{Z}/p \rtimes \langle c^g \rangle$. The second equation indicates that w_1 is an (e'/g)th power of a uniformizer. Thus the third equation indicates that the lower conductor on the generic fibre is $e'g'/g = e/g$. Thus the cover of the generic fibre has inertia group $I_K \simeq \mathbb{Z}/p \rtimes \langle c^g \rangle = \mathbb{Z}/p \rtimes \mu_{m/g}$ and has conductor $e/g = (j_1 r_1 + j_2 r_2)/g$. $\square$

It is more difficult to deform the covers ϕ_1 and ϕ_2 together if p divides $(j_1 r_1 + j_2 r_2)/g$. This is done in the following proposition in the case that $m = 1$ and $p | (j_1 + j_2)$. The conductor on the generic fibre will be slightly bigger than $(j_1 r_1 + j_2 r_2)/g$.

Proposition 2.3.5. *Let $\phi_1 : X \to U$, $\phi_2 : Y \to V$ be $\mathbb{Z}/p$-Galois covers of normal connected germs of curves with conductors j_1 and j_2 respectively. Suppose that $p | (j_1 + j_2)$. Let $e = j_1 + j_2 + \epsilon$ where $\epsilon = 2$ if $p \neq 2$ and $\epsilon = 3$ if $p = 2$. Then there exists a $\mathbb{Z}/p$-Galois cover $\phi_R : W_R \to S^e \simeq S^e_{uv}$ of irreducible germs of R-curves such that:*

1. *The branch locus of the cover ϕ_R consists of exactly one R-point, denoted b_R.*

2. *After normalization, the pullbacks of the special fibre of ϕ_R to U and V, namely $\iota_u^* i^* \phi_k$ and $\iota_v^* i^* \phi_k$, are isomorphic to ϕ_1 and ϕ_2 away from the branch point.*
3. *The generic fibre $\phi_K : X_K \to S_K^e = S^e \times_R K$ of ϕ_R is an I-Galois cover of smooth irreducible germs of curves whose branch locus consists of the point $b_K = b_R \times_R K$.*
4. *The cover ϕ_K has inertia $\mathbb{Z}/p$ and conductor e over the unique branch point b_K.*

Proof. The proof is essentially the same as for Proposition 2.3.4. For $p \neq 2$ one can deform the equations $x^p - x = u^{-j_1}$ and $y^p - y = v^{-j_2}$ using

$$w = u^{j_1+1} + a_1 v^{j_2+1} + d_0 t,\ z^p - z = (u + v + d_1 t)/w.$$

(Here $a_1 = a^m$.) These equations and the Galois action reduce correctly on the components of the special fibre. In particular,

$$\mathrm{mod}(v,t) : w_1 \mapsto u_1^{j_1+1},\ z \mapsto x; \mathrm{mod}(u,t) : w_1 \mapsto a v_1^{j_2+1},\ z \mapsto y/a.$$

Since $uv = t^e = t^{j_1+j_2+2}$, for the same choice of d_0 as in the proof of Proposition 2.3.4, the equations can be rewritten as:

$$w = u^{-(j_2+1)}(u + t^{j_2+1})^{j_1+j_2+2},\ z^p - z = (u + v + d_1 t)/w.$$

Since $p \nmid e$, the conductor is equal to e. □

Notation 2.3.6. Let G be a quasi-p group with Sylow p-subgroup S. Assume $S \simeq \mathbb{Z}/p$. Let $I \simeq \mathbb{Z}/p \rtimes \mu_m \subset G$ and let $r = |G|/mp$. Let $\phi_1 : X \to \mathbb{P}^1_k$ and $\phi_2 : Y \to \mathbb{P}^1_k$ be two (possibly disconnected) G-Galois covers each branched at only one point. Let u (respectively v) be a local parameter at the branch point of ϕ_1 (respectively ϕ_2). Suppose the cover ϕ_i has inertia $I_i \simeq \mathbb{Z}/p \rtimes \mu_{m_i} \subset I$ and conductor j_i for $i = 1, 2$ above $u = 0$ and $v = 0$ respectively. Let the genus of X and Y be g_1 and g_2 respectively.

Let $P_R^{e'}$ be an R-curve whose generic fibre is isomorphic to $\mathbb{P}^1_K$, whose special fibre consists of two projective lines P_u and P_v meeting transversally at a point b where $(u, v) = (0, 0)$, and which satisfies $\hat{P}_{R,b} \simeq S_{uv}^{e'}$.

The next theorem uses Propositon 2.2.1 and Theorem 1.2.4 to deform the covers ϕ_1 and ϕ_2 to a family of covers ϕ_R of $P_R^{e'}$ branched at only one R-point. This family can be defined over a variety Θ of finite type over k. We then specialize to a fibre of the family over another k-point of Θ to get a cover ϕ' with new ramification data.

Theorem 2.3.7. *Consider the pair (ϕ_1, ϕ_2) as in Notation 2.3.2 and 2.3.6. Suppose that ϕ_1 and ϕ_2 satisfy the numerical hypotheses. Then there exist G-Galois covers of curves $\phi_R : Y_R \to P_R^{e'}$ and $\phi' : Y' \to \mathbb{P}^1_k$ such that:*

1. *After normalization, the pullbacks of the special fibre of ϕ_R to P_u and P_v are isomorphic respectively to ϕ_1 and ϕ_2 away from b.*

2. *The branch locus of the cover* ϕ_R *(respectively* ϕ'*) consists of exactly one* R*-point denoted* b_R *which specializes to* b *(respectively of exactly one point* b'*).*
3. *There are* g *ramification points of* ϕ_K *(respectively* ϕ'*) above the branch point* b_K *(respectively above* b'*). These have inertia group* $I_K \simeq \mathbb{Z}/p \rtimes \langle c^g \rangle = \mathbb{Z}/p \rtimes \mu_{m/g}$ *and conductor* $e/g = (j_1 r_1 + j_2 r_2)/g$*.*
4. *The curves* Y_K *and* Y' *are smooth of genus* $g_1 + g_2 - 1 + |G| - r(pm + g - r_1 - r_2)/2$*.*
5. *Suppose* G_1 *and* G_2 *are the stabilizers of a connected component of* X *and* Y *respectively. Then* Y_R *and* Y' *are connected if* G_1 *and* G_2 *generate* G*.*

Proof. After doing the appropriate set-up, the proof is identical to that of Theorem 2.2.2. Let $X^* = P_R^{e'}$ and let $\mathbb{S} = \{b\}$. Let $S \simeq \mathbb{Z}/p$ be a chosen Sylow p-subgroup of G. Then there exist points $\eta_1 \in \phi_1^{-1}(u)$ and $\eta_2 \in \phi_2^{-1}(v)$ with inertia groups $I_i \simeq S \rtimes \mu_{m_i}$. Consider the I_i-Galois covers of germs of curves $\hat{\phi}_1 : \hat{X}_{\eta_1} \to U$ and $\hat{\phi}_2 : \hat{Y}_{\eta_2} \to V$.

Applying Proposition 2.3.4 to the pair $(\hat{\phi}_1, \hat{\phi}_2)$ we see there exists an I-Galois cover $\hat{\phi}_R : W_R \to S^{e'}$ with all the desired properties at the unique branch point b_K. Namely, there are g ramification points above the branch point over K. The inertia group at one of these points is of the form $I_K \simeq \mathbb{Z}/p \rtimes \langle c^g \rangle$ and has conductor e/g. Consider the inclusion $R_s \to A_s$ of rings corresponding to the disconnected G-Galois cover $\mathrm{Ind}_I^G(\hat{\phi}_R)$.

Consider the cover ϕ_k of the special fibre of $P_R^{e'}$ which restricts to ϕ_1 over P_u and to ϕ_2 over P_v. The covers ϕ_k and $\mathrm{Ind}_I^G(\hat{\phi}_R)$ and the isomorphisms given by Proposition 2.3.4 (2) constitute a relative G-Galois thickening problem as in Definition 1.2.1. The (unique) solution to this thickening problem (Theorem 1.2.4) yields the G-Galois cover ϕ_R. Recall from Definition 1.2.3 that the cover ϕ_R is isomorphic to $\mathrm{Ind}_I^G(\hat{\phi}_R)$ over $\hat{P}_{R,b}$. Thus the deformation ϕ_R has the desired properties near the branch point b_R. Also, the cover ϕ_R is isomorphic to the trivial deformation ϕ_{tr} of ϕ away from b, which completes the proof of Properties 1)-3) for ϕ_R. The fact that Y_K is smooth follows because both W_K and ϕ_{tr} are smooth.

If g' is the genus of the fibres of Y_K, the Riemann-Hurwitz formula implies:

$$2g_1 - 2 = -2|G| + rr_1[(pm_1 - 1) + j_1(p-1)],$$
$$2g_2 - 2 = -2|G| + rr_2[(pm_2 - 1) + j_2(p-1)],$$
$$2g' - 2 = -2|G| + rg[(pm/g - 1) + e(p-1)/g].$$

Using the fact that $e = j_1 r_1 + j_2 r_2$ it follows that:

$$g' = g_1 + g_2 - 1 + |G| - r(pm + g - r_1 - r_2)/2.$$

Note that $pm + g - r_1 - r_2$ is always even. Finally, if G_1 and G_2 generate G then the special fibre of Y_R is connected. Thus Y_K is connected.

As in the proof of Theorem 2.2.2, the cover ϕ_R can be defined over Spec(Θ) for some $k \neq \Theta \subset R$ of finite type over k. Also one can choose a fibre ϕ' : $Y' \to \mathbb{P}^1_k$ over a k-point of Θ so that ϕ' is a G-Galois cover and Y' is smooth with the same number of connected components as Y_K, [**2**, Proposition 9.29]. Properties (2)-(5) for ϕ' follow immediately from the corresponding properties for ϕ_K. □

Remark 2.3.8. These patching results do not need to be restricted to the case $X = \mathbb{P}^1_k$ or the case of only one branch point. In general, one can consider G-Galois covers $\phi_1 : Y_1 \to X_1$ and $\phi_2 : Y_2 \to X_2$. Suppose ϕ_1 and ϕ_2 each have a branch point with inertia group contained in $I \simeq \mathbb{Z}/p \rtimes \mu_m$ whose ramification data satisfies the numerical hypotheses. Using Proposition 2.3.4, one can construct a cover $\phi : Z \to W$ with specified inertia behavior above one point. The genus of W will be the sum of genus(X_1) and genus(X_2) and ϕ will be branched at $|B_1| + |B_2| - 1$ points.

Remark 2.3.9. One would like to know whether the above constructions are optimal in the following sense: Given the covers ϕ_1 and ϕ_2, with upper jumps $\sigma_1 = j_1/m_1$ and $\sigma_2 = j_2/m_2$, in Theorem 2.3.7 we construct a deformation so that the generic fibre is a cover ϕ_K branched at exactly one point with upper jump $\sigma_\eta = \sigma_1 + \sigma_2$; would it have been possible to get any smaller upper jump on the generic fibre? The key formula [**7**, Theorem 3.11] indicates that the result in Theorem 2.3.7 is almost optimal since the upper jump on the generic fibre must satisfy $\sigma_\eta \geq \sigma_1 + \sigma_2 - 1$. In other words, the singularity for ϕ_k is not too severe. The following lemma gives another way to measure this singularity.

Consider the cover $\phi_R : Y_R \to \mathbb{P}^1_R$ constructed in Theorem 2.2.2 (respectively $\phi_R : Y_R \to P^{e'}_R$ constructed in Theorem 2.3.7 or Corollary 2.3.1). Choose $y \in \phi_R^{-1}(\infty_k)$ (respectively $\phi_R^{-1}(0,0)$). Let $\pi_y : \widetilde{Y}_{y,k} \to \hat{Y}_{y,k}$ be the normalization of $\hat{Y}_{y,k}$ and let $\hat{\mathcal{O}}_y \to \widetilde{\mathcal{O}}_y$ be the corresponding extension of rings. Let $\delta_y = \dim_k(\widetilde{\mathcal{O}}_y/\mathcal{O}_y)$ and let $m_y = \#\pi_Y^{-1}(y)$. Let $\mu_y = 2\delta_y - m_y + 1$.

Lemma 2.3.10.

i) *In Theorem* 2.2.2*,* $\mu_y = (j_\eta - j_b)(p-1) = im(p-1)$*.*
ii) *In Theorem* 2.3.7 *and Corollary* 2.3.1 (2-3)*,* $\mu_y = 1 + mp - g$*.*
iii) *In Corollary* 2.3.1 (1)*,* $\mu_y = p + \epsilon(p-1)$*.*

Proof. The proof uses a formula of Kato [**6**] which compares the local ramification and the singularities for the cover $\hat{\phi}_R$ of germs of curves. The details are omitted. □

3. Applications to ramification questions.

Let G be a finite quasi-p group. Let S be a chosen Sylow p-subgroup of G and suppose S has order p. Let k be an algebraically closed field of characteristic

p. All covers in this section are smooth, connected and proper. We now show that for all sufficiently large $j \in \mathbb{N}$ with $p \nmid j$, there exists a G-Galois cover $\phi : Y \to \mathbb{P}^1_k$ branched at only one point with inertia $\mathbb{Z}/p$ and conductor j. The method is to first show the existence of such a cover with small conductor under an additional hypothesis on G. We then use group theory to determine which conductors are sufficiently large enough to realize with formal patching.

3.1. *P*-Weight. In this section, we measure the complexity of the group G.

Definition 3.1.1. Let $G(S) \subset G$ be the subgroup generated by all proper quasi-p subgroups G' such that $G' \cap S$ is a Sylow p-subgroup of G'. The group G is *p-pure* if $G(S) \neq G$.

Note that this definition is independent of the choice of S. This condition was introduced in [**10**]. Note that when $|S| = p$, then G is p-pure if and only if G is not generated by all proper quasi-p subgroups $G' \subset G$ such that $S \subset G'$. Some examples of p-pure quasi-p groups with $|S| = p$ are $\mathrm{PSL}_2(\mathbb{F}_p)$, and the semi-direct product $(\mathbb{Z}/r\mathbb{Z})^l \rtimes \mathbb{Z}/p$ where the action is irreducible. When $p = 11$, M_{11} and M_{22} are quasi-11 and 11-pure. Every finite quasi-p group can be generated from p-pure ones.

Definition 3.1.2. Consider all subgroups $G' \subset G$ such that G' is quasi-p and p-pure and such that $G' \cap S$ is a Sylow p-subgroup of G'. The *p-weight* $\omega(G)$ of G is the minimal number of such subgroups G' of G which are needed to generate G.

Lemma 3.1.3. *The p-weight $\omega(G)$ of G is a finite number independent of the choice of S.*

Proof. The proof uses induction on $|G|$ to show that G can be generated by p-pure quasi-p subgroups G'' with $G'' \cap S = \mathrm{Syl}_p(G'')$. This statement is true if $G \simeq \mathbb{Z}/p$. For then G contains no proper quasi-p subgroups and so $\{1\} = G(S) \neq G$. Thus $G \simeq \mathbb{Z}/p$ is p-pure and $\omega(G) = 1$.

Now given G, suppose that the hypothesis is true for any quasi-p group G' such that $p \leq |G'| < |G|$. If $G(S) \neq G$ then G is p-pure and so $\omega(G) = 1$.

If $G(S) = G$ then by definition G is generated by its proper quasi-p subgroups $G' \subset G$ with $G' \cap S = \mathrm{Syl}_p(G')$. Since $|G'| < |G|$ the induction hypothesis states that each G' is generated by p-pure quasi-p subgroups G'' with $G'' \cap \mathrm{Syl}_p(G') = \mathrm{Syl}_p(G'')$. Note that $\mathrm{Syl}_p(G'') = G'' \cap (G' \cap S) = G'' \cap S$. Thus each G'' satisfies the necessary conditions and the collection of G'' generate G. Thus the p-weight is the minimum size among all sets of p-pure quasi-p subgroups G'' with $S \cap G'' = \mathrm{Syl}_p(G'')$ which generate G.

To show that $\omega(G)$ is independent of S, consider another Sylow p-subgroup S_0 of G. Let $\omega_0(G)$ be the p-weight with respect to S_0. Since the Sylow

p-subgroups are all conjugate, there exists some $g \in G$ with $S_0 = gSg^{-1}$. Suppose G is generated by a set $\{G''\}$ of p-pure quasi-p subgroups with $S \cap G'' = \mathrm{Syl}_p(G'')$. Note that $S_0 \cap (gG''g^{-1}) = \mathrm{Syl}_p(gG''g^{-1})$. Each subgroup $gG''g^{-1}$ is still quasi-p and p-pure (with respect to its Sylow). Also G is generated by the set of $gG''g^{-1}$. Thus $\omega_0(G) \leq \omega(G)$. Reversing the roles of S and S_0, $\omega(G) \leq \omega_0(G)$. □

3.2. Conductors. Let $\phi : Y \to \mathbb{P}^1_k$ be a G-Galois cover which is branched at only one point. Such a cover exists if and only if G is a quasi-p group which means that G is generated by p-groups, [**10**]. Suppose ϕ has inertia group $I \simeq \mathbb{Z}/p \rtimes \mu_m$ and conductor j. When $G \neq \mathbb{Z}/p$, there is a small set of values $j_{\min}(I)$, depending only on I, consisting of the minimal possible conductors for ϕ. Let n' be the order of the prime-to-p part of the center of I. Let n be such that $m = nn'$.

Definition 3.2.1. Define $j_{\min}(I) = \{j_{\min}(I,a) |\ 1 \leq a \leq n,\ \gcd(a,n) = 1\}$ where $j_{\min}(I,a) = 2m + n'$ if $a = 1$ and $n = p-1$ and $j_{\min}(I,a) = m + an'$ otherwise.

A geometric interpretation for the set $j_{\min}(I)$ is that ϕ has a non-isotrivial deformation in equal characteristic p if and only if $j \notin j_{\min}(I)$, [**8**, Theorem 3.1.11]. Suppose $1 \leq a \leq n$ and $j \equiv an' \bmod m$. If $G \neq \mathbb{Z}/p$ then $j \geq j_{\min}(I,a)$, by [**8**, Lemma 1.4.3]. Note that if $j \in j_{\min}(I)$ then $p \nmid j$ and $j \leq m(2 + 1/(p-1))$.

Theorem 3.2.2. *Let G be a finite p-pure quasi-p group whose Sylow p-subgroups have order $p \neq 2$. For some $I \simeq \mathbb{Z}/p \rtimes \mu_m \subset G$ and some $j \in j_{\min}(I)$, there exists a G-Galois cover $\phi : Y \to \mathbb{P}^1_k$ of smooth connected curves branched at only one point over which it has inertia group I and conductor j. In particular,* $\mathrm{genus}(Y) \leq 1 + |G|(p-1)/2p$.

Proof. For the convenience of the reader we briefly recall the outline of the proof from [**7**, Theorem 4.5]. By Abhyankar's Conjecture [**10**, 6.5.3], for some I_0 of the form $\mathbb{Z}/p \rtimes \mu_{m_0}$ and some j_0, there exists a G-Galois cover $\phi_0 : Y_0 \to \mathbb{P}^1_k$ with group G which is branched at only one point with inertia group I_0 and conductor j_0. If $j_0 \notin j_{\min}(I)$, there exists a non-isotrivial deformation of ϕ_0 in equal characteristic p by [**8**, Theorem 3.1.11]. This deformation yields a cover ϕ_K with bad reduction by [**8**, Theorem 3.3.7].

Let $\phi : Y \to X$ be the stable model of ϕ_K. See [**7**, Section 3] for information on the structure of ϕ, which is very similar to that in the unequal characteristic case in [**10**, Section 6], [**11**, Sections 2-3], and [**12**]. In particular, the special fibre X_k is a tree of projective lines and the restriction of ϕ over any terminal component of X_k is separable. Since G is p-pure and has no (nontrivial) normal p-subgroups, for some terminal component P_b of X_k, the curve $Y_b = \phi^{-1}(P_b)$ is connected. For this component, the restriction $\phi_b : Y_b \to P_b \simeq \mathbb{P}^1_k$ is a G-Galois cover branched at only one point. If ϕ_b has inertia group $I_b \simeq \mathbb{Z}/p \rtimes \mu_{m_b} \subset N_G(S)$ and conductor j_b then $j_b/m_b < j_0/m_0$ by [**7**, Theorem 3.11]. We reiterate this process until finding such a cover with inertia group $I_b = \mathbb{Z}/p \rtimes \mu_{m_b}$ and conductor j_b satisfying $j_b/m_b \leq 2 + 1/(p-1)$ which implies $j_b \in j_{\min}(I)$. The condition on genus(Y) follows directly from Definition 3.2.1 and the Riemann-Hurwitz formula. □

Lemma 3.2.3. *Suppose there exists a G-Galois cover $\phi : Y \to \mathbb{P}^1_k$ branched at only one point with inertia group $I \simeq \mathbb{Z}/p \rtimes \mathbb{Z}/m$ and conductor j. Then for any $I' \subset I$ of the form $I' = \mathbb{Z}/p \rtimes \mu_{m'}$ and for any $j' \equiv j \bmod m'$ with $j' \geq j$ and $p \nmid j'$, there exists a G-Galois cover $\phi' : Y \to \mathbb{P}^1_k$ of smooth connected curves branched at only one point with inertia I' and conductor j'.*

Proof. Let r be the index of I' in I. Consider the cover $f : X \to \mathbb{P}^1_k$ which is cyclic of order r and branched at 0 and ∞. By Abhyankar's Lemma, the cover $f^*\phi$ is a G-Galois cover $\phi' : Y' \to \mathbb{P}^1_k$ which is branched at only one point with inertia group $\mathbb{Z}/p \rtimes \mu_{m'}$. The cover $f^*\phi$ is connected since f and ϕ are disjoint. The conductor of $f^*\phi$ still equals j. Thus the statement is immediate from Theorem 2.2.2. □

Let G be a quasi-p group with $|S| = p$ and with p-weight ω. By Lemma 3.1.3, G can be generated by a collection of ω proper p-pure quasi-p subgroups G' such that $G' \cap S = \mathrm{Syl}_p(G')$. We give sufficient conditions for the conductor of a G-Galois cover $\phi : Y \to \mathbb{P}^1_k$ branched at only one point with inertia $\mathbb{Z}/p$.

Theorem 3.2.4. *Let G be a finite quasi-p group whose Sylow p-subgroups have order $p \neq 2$. Let ω be the p-weight of G. Let m_e be the exponent of the normalizer $N_G(S)$ of S in G divided by p. Let $j \in \mathbb{N}^+$ satisfy $\gcd(j,p) = 1$. Suppose $j \geq m_e(2 + 1/(p-1))\omega$ if $p \nmid \omega$ and $j \geq m_e(2 + 1/(p-1))\omega + 2$ if $p|\omega$. Then there exists a G-Galois cover $\phi : Y \to \mathbb{P}^1_k$ of smooth connected curves which is branched at only one point over which it has inertia $\mathbb{Z}/p$ and conductor j.*

Proof. Note that $m_e(2 + 1/(p-1))$ is not necessarily an integer. In this proof the phrase "cover of this type" indicates that the cover in question is a smooth connected cover of the projective line branched at only one point with inertia $I = \mathbb{Z}/p$.

By Theorem 2.2.2, given G as above it is sufficient to prove the following: For some $J \in \mathbb{Z}$ such that $p \nmid J$ and $J \leq m_e(2 + 1/(p-1))$ there exists a G-Galois cover ϕ of this type with conductor $j = J\omega$ if $p \nmid \omega$ and conductor $j = J\omega + 2$ if $p|\omega$. The proof will proceed by induction on ω.

If $\omega = 1$ then G is quasi-p and p-pure. By Theorem 3.2.2, for some $I \simeq \mathbb{Z}/p \rtimes \mu_m \subset G$ there exists a G-Galois cover $\phi : Y \to \mathbb{P}^1_k$ branched at only one point with inertia group I and conductor $j \in j_{\min}(I)$. Recall that if $j \in j_{\min}(I)$ then $j = m(2 + 1/(p-1))$ or $j \leq 2m$. Also $m \leq m_e$. Let $I' = \mathbb{Z}/p$. By Lemma 3.2.3, there exists a G-Galois cover $\phi' : Y \to \mathbb{P}^1_k$ branched at only one point with inertia $\mathbb{Z}/p$ and conductor j. Choose $J = j$ and note that $p \nmid J$ and $J \leq m_e(2 + 1/(p-1))$.

Now suppose that $\omega > 1$. By the inductive hypothesis, for all quasi-p groups G' having p-weight ω' where $\omega' < \omega$ and $p \nmid \omega'$, there exists J' such that $\gcd(J', p) = 1$ and such that $J' \leq m_{e'}(2 + 1/(p-1))$ (where $m_{e'}$ is the exponent of $N_{G'}(S)$ divided by p) and there exists a G'-Galois cover ϕ' of this type with conductor $j' = J'\omega'$.

Choose $w_1 \geq 1$ and $w_2 \geq 1$ such that $p \nmid w_1 w_2$ and $w_1 + w_2 = \omega$. (If $\omega \not\equiv 1 \bmod p$, then choose $w_1 = 1$ and $w_2 = \omega - 1$. If $\omega \equiv 1 \bmod p$, then choose $w_1 = 2$ and $w_2 = \omega - 2$.)

Since G has p-weight $\omega > 1$, G can be generated by ω proper p-pure quasi-p groups $G'_1, \ldots, G'_\omega$ with $S \subset G'_i$ for all $1 \leq i \leq \omega$. Let $G_1 \subset G$ be the subgroup generated by G'_i for $1 \leq i \leq w_1$. Let $G_2 \subset G$ be the subgroup generated by G'_i for $w_1 + 1 \leq i \leq w_2$. Then G_1 and G_2 are quasi-p groups since they are generated by quasi-p groups. For $i = 1, 2$, the order of a Sylow p-subgroup of G_i is p since $S \subset G_i \subset G$. By construction, $\omega(G_1) \leq w_1$ and $\omega(G_2) \leq w_2$. But in fact, $\omega(G_1) = w_1$ and $\omega(G_2) = w_2$ since the p-weight of G is only ω. For $i = 1, 2$, let m_{e_i} be the exponent of $N_{G_i}(S)$ divided by p.

By the inductive hypothesis, for $i = 1, 2$, there exists J_i such that $p \nmid J_i$ and $J_i \leq m_{e_i}(2 + 1/(p-1))$ and there exists a G_i-Galois cover ϕ_i of this type with conductor $J_i w_i$. Let m_e be the exponent of the normalizer of $N_G(S)$ divided by p and note that $m_e \geq m_{e_i}$. Let $J = \max\{J_1, J_2\}$ and note that $p \nmid J$ and $J \leq m_e(2 + 1/(p-1))$. Using Theorem 2.2.2 to increase the conductor of ϕ_i for $i = 1, 2$ we find a G_i-Galois cover ϕ'_i of this type with conductor Jw_i.

By Corollary 2.3.1 (1), in the case that $p \nmid \omega$, there exists a G-Galois cover ϕ of this type with conductor $j = Jw_1 + Jw_2 = J\omega$; in the case that $p | \omega$, there exists a G-Galois cover ϕ of this type with conductor $j = Jw_1 + Jw_2 + 2 = J\omega + 2$. (In particular, the covers are connected since G_1 and G_2 generate G, Theorem 2.3.7 (5).) This completes the proof by induction. $\square$

Example 3.2.5. Let $G = \mathrm{PSL}_2(\mathbb{F}_p)$. Then G is quasi-p and p-pure and its Sylow p-subgroups have order p. The normalizer of a Sylow is of the form $I^* \simeq \mathbb{Z}/p \rtimes \mu_{m^*}$ where $m^* = (p-1)/2$ and μ_{m^*} acts faithfully on $\mathbb{Z}/p$. By Corollary 3.2.2, for some $m | m^*$, there exists a G-Galois cover $\phi : Y \to \mathbb{P}^1_k$ of smooth connected curves branched only at ∞ with inertia $I \simeq \mathbb{Z}/p \rtimes \mu_m$ and conductor $j \in j_{\min}(I)$. In this case $j \leq 2m \leq 2m^* < p$.

Acknowledgements. I would like to thank D. Harbater and K. Stevenson for their help improving earlier drafts of this work.

References

[1] I. Bouw and R. Pries, *Rigidity, reduction, and ramification*, to appear in Math. Ann.

[2] M. Fried and M. Jarden, *Field Arithmetic*, Springer-Verlag, Berlin, 1986, MR 89b:12010, Zbl 0625.12001.

[3] D. Harbater, *Formal patching and adding branch points*, Amer. J. Math., **115**(**3**) (1993), 487-508, MR 94f:14024, Zbl 0790.14027.

[4] ———, *Abhyankar's conjecture on Galois groups over curves*, Invent. Math., **117**(**1**) (1994), 1-25, MR 95i:14029, Zbl 0805.14014.

[5] D. Harbater and K. Stevenson, *Patching and thickening problems*, J. Algebra, **212**(**1**) (1999), 272-304, MR 2000b:14004, Zbl 0944.14013.

[6] K. Kato, *Vanishing cycles, ramification of valuations, and class field theory*, Duke Math. J., **55**(**3**) (1987), 629-659, MR 88k:11083, Zbl 0665.14005.

[7] R. Pries, *Conductors of wildly ramified covers*, I, C.R. Acad. Sci. Paris Sér. I Math., **335**(**1**) (2002), 481-484, CMP 1 937 117.

[8] ———, *Families of wildly ramified covers of curves*, Amer. J. Math., **124**(**4**) (2002), 737-768, CMP 1 914 457.

[9] ———, *Construction of covers with formal and rigid geometry*, in 'Courbes semi-stables et groupe fondamental en géométrie algébrique (Luminy, 1998),' 157-167, Birkhäuser, Basel, 2000, MR 2001e:14026, Zbl 0978.14018.

[10] M. Raynaud, *Revêtements de la droite affine en caractéristique $p > 0$ et conjecture d'Abhyankar*, Invent. Math., **116**(**1-3**) (1994), 425-462, MR 94m:14034, Zbl 0798.14013.

[11] ———, *Spécialisation des revêtements en caractéristique $p > 0$*, Ann. Sci. École Norm. Sup. (4), **32**(**1**) (1999), 87-126, MR 2000e:14016, Zbl 0999.14004.

[12] M. Saïdi, *Raynaud's proof of Abhyankar's conjecture for the affine line*, in 'Courbes semi-stables et groupe fondamental en géométrie algébrique,' Progress in Mathematics, **187**, 249-265, Birkhäuser, 2000.

[13] J.-P. Serre, *Corps Locaux*, Hermann, 1968, MR 50 #7096, Zbl 0174.24301.

Received June 13, 2002 and revised November 2, 2002.

Department of Mathematics
Columbia University
New York, NY 10027
E-mail address: pries@math.columbia.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 211, No. 1, 2003

LOCAL MOVES ON SPATIAL GRAPHS AND FINITE TYPE INVARIANTS

Kouki Taniyama and Akira Yasuhara

We define A_k-moves for embeddings of a finite graph into the 3-sphere for each natural number k. Let A_k-equivalence denote an equivalence relation generated by A_k-moves and ambient isotopy. A_k-equivalence implies A_{k-1}-equivalence. Let $\mathcal{F}$ be an A_{k-1}-equivalence class of the embeddings of a finite graph into the 3-sphere. Let $\mathcal{G}$ be the quotient set of $\mathcal{F}$ under A_k-equivalence. We show that the set $\mathcal{G}$ forms an abelian group under a certain geometric operation. We define finite type invariants on $\mathcal{F}$ of order $(n;k)$. And we show that if any finite type invariant of order $(1;k)$ takes the same value on two elements of $\mathcal{F}$, then they are A_k-equivalent. A_k-move is a generalization of C_k-move defined by K. Habiro. Habiro showed that two oriented knots are the same up to C_k-move and ambient isotopy if and only if any Vassiliev invariant of order $\leq k-1$ takes the same value on them. The 'if' part does not hold for two-component links. Our result gives a sufficient condition for spatial graphs to be C_k-equivalent.

Introduction.

K. Habiro defined a local move, *C_k-move*, for each natural number k [**2**]. It is known that if two embeddings f and g of a graph into the three sphere are the same up to C_k-move and ambient isotopy, then g can be deformed into a band sum of f with certain $(k+1)$-component links and that changing position of a band and an arc, which is called *a band trivialization* of C_k-move, is realized by C_{k+1}-moves and ambient isotopy [**17**]. This is one of the most important properties of C_k-move. We consider local moves which have this property. We define A_1-move as the crossing change and A_{k+1}-move as a band trivialization of A_k-move; see Section 1 for the precise definition. So A_k-move is a generalization of C_k-move. In fact, the results for A_k-move in this paper hold for C_k-move.

Let *A_k-equivalence* denote an equivalence relation given by A_k-moves and ambient isotopy. Habiro showed that two oriented knots are C_k-equivalent if and only if they have the same Vassiliev invariants of order $\leq k-1$ [**3**], [**4**]. The 'only if' part of this result is true for A_k-move and for the embeddings of a graph, in particular for links (Theorem 5.1). However the 'if' part does not hold for two-component links. For example, the Whitehead link is not C_3-equivalent to a trivial link because they have different Arf invariants, see [**16**]. On the other hand, H. Murakami showed in [**7**] that the Vassiliev invariants of links of order ≤ 2 are determined by the linking numbers and the second coefficient of the Conway polynomial of each component. Hence, the values of any Vassiliev invariant of order ≤ 2 of these two links are the same. So we note that Vassiliev invariants of order $\leq k-1$ are not enough to characterize C_k-equivalent embeddings of a graph.

We will define in Section 1 a *finite type invariant of order* $(n; k)$ as a generalization of a Vassiliev invariant and see that if any finite type invariants of order $(1; k)$ takes the same value on two A_{k-1}-equivalent embeddings of a graph, then they are A_k-equivalent (Theorem 1.1). While a Vassiliev invariant is defined by the change in its value at every 'wall' corresponding to a crossing change, a finite type invariant of order $(n; k)$ is defined similarly by 'walls' corresponding to A_k-moves. A finite type invariant of order $(n; 1)$ is a Vassiliev invariant of order $\leq n$.

It is shown that the set of C_k-equivalence classes of knots forms an abelian group under the connected sum [**3**], [**4**]. This is also true for A_k-equivalence classes. Since the connected sum is peculiar to knots, we cannot apply it to embeddings of a graph. In Section 2, we will define a certain geometric sum for the elements in an A_{k-1}-equivalence class of the embedding of a graph. Then we will see that the quotient set of the A_{k-1}-equivalence class under A_k-equivalence forms an abelian group (Theorem 2.4).

It is not essential that A_1-move is the crossing change. This is a big difference between A_k-move and C_k-move. We will study a generalization of A_k-move in Section 4. For example, if we put A_1-move to be the #-move defined by Murakami [**6**], then we get several results similar to that for original A_k-move.

1. A_k-moves and finite type invariants.

Let B^3 be the oriented unit 3-ball. A *tangle* is a disjoint union of properly embedded arcs in B^3. A tangle is *trivial* if it is contained in a properly embedded 2-disk in B^3. A *trivialization* of a tangle $T = t_1 \cup t_2 \cup \cdots \cup t_k$ is a choice of mutually disjoint disks $D_1, D_2, \ldots, D_k$ in B^3 such that $D_i = (D_i \cap \partial B^3) \cup t_i$ for $i = 1, 2, \ldots, k$. It can be shown that in general a trivialization is not unique up to ambient isotopy of B^3 fixed on the tangle.

Let T and S be tangles, and let $t_1, t_2, \ldots, t_k$ and $s_1, s_2, \ldots, s_k$ be the components of T and S respectively. Suppose that for each t_i there exists some s_j such that $\partial t_i = \partial s_j$. Then we call the ordered pair (T, S) a *local move*, which can be interpreted as substituting S for T. Two local moves (T, S) and (T', S') are *equivalent* if there exists an orientation preserving homeomorphism $h : B^3 \longrightarrow B^3$ such that $h(T) = T'$ and $h(S)$ is ambient isotopic to S' relative to ∂B^3. We consider local moves up to this equivalence.

Let (T, S) be a local move such that T and S are trivial tangles. First choose a trivialization $D_1, D_2, \ldots, D_k$ of T. Each D_i intersects ∂B^3 in an arc γ_i. Let E_i be a small regular neighbourhood of γ_i in ∂B^3. We devide the circle ∂E_i into two arcs α_i and β_i such that $\alpha_i \cap \beta_i = \partial\alpha_i = \partial\beta_i$. By slightly perturbing intα_i and intβ_i into the interior of B^3 on either side of D_i, we obtain properly embedded arcs $\widetilde{\alpha}_i$ and $\widetilde{\beta}_i$. We consider k local moves $(S \cup \widetilde{\alpha}_i, S \cup \widetilde{\beta}_i)$ $(i = 1, 2, \ldots, k)$ and call them the *band trivializations* of the local move (T, S) with respect to the trivialization $D_1, D_2, \ldots, D_k$. Note that both $S \cup \widetilde{\alpha}_i$ and $S \cup \widetilde{\beta}_i$ are trivial tangles.

We now inductively define a sequence of local moves on trivial tangles in B^3 which depend on the choice of trivialization. An A_1*-move* is the crossing change shown in Figure 1.1. Suppose that A_k-moves are defined and there are l A_k-moves $(T_1, S_1), (T_2, S_2), \ldots, (T_l, S_l)$ up to equivalence. For each A_k-move (T_i, S_i) $(i = 1, 2, \ldots, l)$, we choose a single trivialization $\tau_i = \{D_{i,1}, D_{i,2}, \ldots, D_{i,k+1}\}$ of T_i and fix it. (The choice of τ_i is independent of the trivialization that is chosen to define A_k-move (T_i, S_i).) Then the band trivializations of (T_i, S_i) with respect to the trivialization τ_i are called $A_{k+1}(\tau_i)$*-moves* and these $A_{k+1}(\tau_i)$-moves $(i = 1, 2, \ldots, l)$ are called $A_{k+1}(\tau_1, \tau_2, \ldots, \tau_l)$-moves. Note that the number of $A_{k+1}(\tau_1, \tau_2, \ldots, \tau_l)$-moves is at most $l(k + 1)$ up to equivalence. Although the choice of trivializations is important for the definition of A_k-move, our proof is the same for every choice. Therefore the results of this paper hold for every choice of trivializations $\tau_1, \tau_2, \ldots, \tau_l$. So we denote $A_{k+1}(\tau_1, \tau_2, \ldots, \tau_l)$-move simply as A_{k+1}*-move.* It is known that C_k-move defined by Habiro is a special case of A_k-move for certain choices of trivializations; see [**2**], [**10**]. We will see that A_k-move, as well as C_k-move, has the property mentioned in Introduction (Proposition 2.1 and Lemma 2.2).

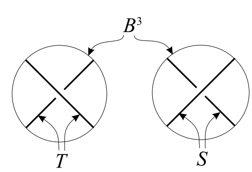
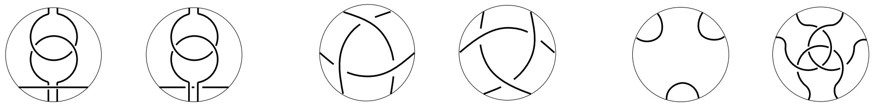


Figure 1.1.

Examples. (1) The trivialization of a tangle in Figure 1.1 is unique up to ambient isotopy. Therefore we have any band trivialization of an A_1-move is equivalent to the local move in Figure 1.2-(i). Thus A_2-move is unique up to equivalence. It is not hard to see that an A_2-move is equivalent to the *delta move* in Figure 1.2-(ii) defined by H. Murakami and Y. Nakanishi [**8**], and then it is equivalent to the local move in Figure 1.2-(iii).

(2) If we choose a trivialization for the A_2-move as in Figure 1.3-(i), then, by the symmetry of the A_2-move, any A_3-move is equivalent to the local move in Figure 1.3-(ii).



(i) (ii) (iii)

Figure 1.2.

(i) (ii)

Figure 1.3.

A local move (S, T) is called the *inverse* of a local move (T, S). It is clear that the inverse of an A_1-move is again an A_1-move. By the definition of A_k-move, we see that the inverse of an A_k-move with $k \geq 2$ is equivalent to itself.

Let (T, S) be an A_k-move and $D_1, D_2, \ldots, D_{k+1}$ the fixed trivialization of $T = t_1 \cup t_2 \cup \cdots \cup t_{k+1}$. We set $\alpha = \partial B^3 \cap (D_1 \cup D_2 \cup \cdots \cup D_{k+1})$ and $\beta = S$. A link L in S^3 is called *type* k if there is an orientation preserving embedding $\varphi : B^3 \longrightarrow S^3$ such that $L = \varphi(\alpha \cup \beta)$. Then the pair (α, β) is called a *link model of* L.

We now define an equivalence relation on spatial graphs by A_k-move. Let G be a finite graph. Let $V(G)$ denote the set of the vertices of G. Let $f, g : G \longrightarrow S^3$ be embeddings. We say that f and g are *related by an* A_k-*move* if there is an A_k-move (T, S) and an orientation preserving embedding $\varphi : B^3 \longrightarrow S^3$ such that:

(i) If $f(x) \neq g(x)$ then both $f(x)$ and $g(x)$ are contained in $\varphi(\text{int}B^3)$,
(ii) $f(V(G)) = g(V(G))$ is disjoint from $\varphi(B^3)$, and
(iii) $f(G) \cap \varphi(B^3) = \varphi(T)$ and $g(G) \cap \varphi(B^3) = \varphi(S)$.

We also say that g is obtained from f by an *application* of (T, S). We define A_k-*equivalence* as an equivalence relation on the set of all embeddings of G into S^3 given by the relation above and ambient isotopy. For an embedding $f : G \longrightarrow S^3$, let $[f]_k$ denote the A_k-equivalence class of f. By the definition of A_k-move we see that an application of an A_{k+1}-move is realized by two applications of A_k-move and ambient isotopy. Thus A_{k+1}-equivalence implies A_k-equivalence. In other words we have $[f]_1 \supset [f]_2 \supset \cdots \supset [f]_k \supset [f]_{k+1} \supset \cdots$.

Let $f : G \longrightarrow S^3$ be an embedding, L_i links of type k and (α_i, β_i) their link models ($i = 1, 2, \ldots, n$). Let $I = [0, 1]$ be the unit closed interval. An embedding $g : G \longrightarrow S^3$ is called a *band sum of f with $L_1, L_2, \ldots, L_n$* if there are mutually disjoint embeddings $b_{ij} : I \times I \longrightarrow S^3$ ($i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, k+1$) and mutually disjoint orientation preserving embeddings $\varphi_i : B^3 \longrightarrow S^3 - f(G)$ with $L_i = \varphi_i(\alpha_i \cup \beta_i)$ ($i = 1, 2, \ldots, n$) such that the following (i) and (ii) hold:

(i) $b_{ij}(I \times I) \cap f(G) = b_{ij}(I \times I) \cap f(G - V(G)) = b_{ij}(I \times \{0\})$ and $b_{ij}(I \times I) \cap (\bigcup_l \varphi_l(B^3)) = b_{ij}(I \times \{1\})$ is a component of $\varphi_i(\alpha_i)$ for any i, j ($i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, k+1$).

(ii) $f(x) = g(x)$ if $f(x)$ is not contained in $\bigcup_{i,j} b_{ij}(I \times \{0\})$ and

$$g(G) = \left(f(G) \cup \bigcup_i L_i - \bigcup_{i,j} b_{ij}(I \times \partial I) \right) \cup \bigcup_{i,j} b_{ij}(\partial I \times I).$$

Then we denote g by $F(f; \{L_1, L_2, \ldots, L_n\}, \{B_1, B_2, \ldots, B_n\})$, where $B_i = b_{i1}(I \times I) \cup b_{i2}(I \times I) \cup \cdots \cup b_{ik+1}(I \times I)$ ($i = 1, 2, \ldots, n$). We call each $b_{ij}(I \times I)$ a *band*. We call each $\varphi_i(B^3)$ an *associated ball* of L_i. See Figure 1.4 for an example of a band sum of an embedding f with links L_1, L_2, L_3 of type 3.

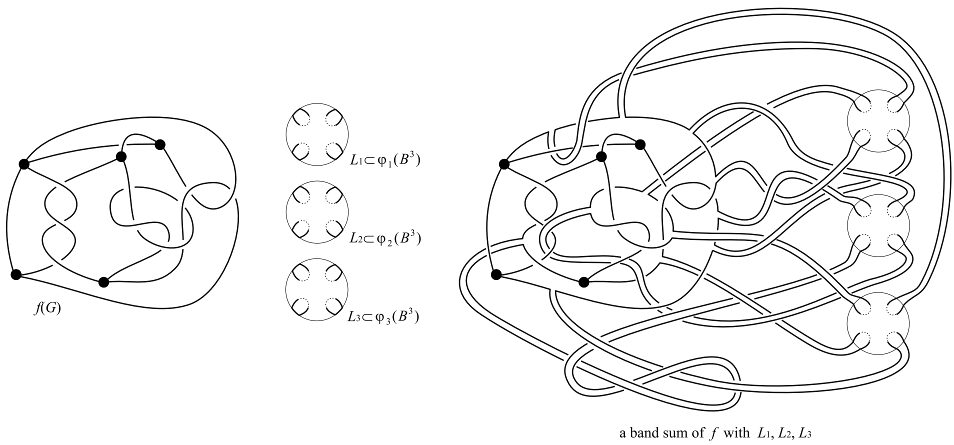


Figure 1.4.

Remark. It follows from the definition that if g is a band sum of f with some links of type k, then g is A_k-equivalent to f. The converse is also true and will be shown in Proposition 2.1. In Lemma 2.2, we show that the position of a band is changeable up to A_{k+1}-equivalence. The origin of the name 'band trivialization' comes from this fact.

Let $h : G \longrightarrow S^3$ be an embedding and H an abelian group. Let $\varphi : [h]_{k-1} \longrightarrow H$ be an invariant. We say that φ is a *finite type invariant of order* $(n; k)$ if for any embedding $f \in [h]_{k-1}$ and any band sum $F(f;\{L_1,L_2,\dots,L_{n+1}\},\{B_1,B_2,\dots,B_{n+1}\})$ of f with links $L_1,L_2,\dots,L_{n+1}$ of type $k-1$,

$$\sum_{X\subset\{1,2,\dots,n+1\}}(-1)^{|X|}\varphi\left(F\left(f;\bigcup_{i\in X}\{L_i\},\bigcup_{i\in X}\{B_i\}\right)\right)=0\in H,$$

where the sum is taken over all subsets, including the empty set, and $|X|$ is the number of the elements in X.

In the next section we show the following theorem:

Theorem 1.1. *Let $f,g:G\longrightarrow S^3$ be A_{k-1}-equivalent embeddings. Then they are A_k-equivalent if and only if $\varphi(f)=\varphi(g)$ for any finite type A_k-equivalence invariant φ of order $(1;k)$.*

Note that finite type invariants of order $(n;2)$ coincide with Vassiliev invariants of order n. It is shown in [**5**, Theorem 1.1, Theorem 1.3] that two embeddings of a finite graph G into S^3 are A_2-equivalent if and only if they have the same Wu invariant [**18**]. It follows from [**14**, Section 2] that Wu invariant is a finite type invariant of order $(1;2)$. Since two embeddings are always A_1-equivalent, we have the following corollary:

Corollary 1.2. *Let $f,g:G\longrightarrow S^3$ be embeddings. Then the following conditions are mutually equivalent:*

(i) *f and g are A_2-equivalent.*
(ii) *f and g have the same Wu invariant.*
(iii) *$\varphi(f)=\varphi(g)$ for any Vassiliev invariant φ of order* 1.

In Section 5 we show the following proposition:

Proposition 1.3. *Let φ be a Vassiliev invariant of order $(n+1)(k-1)-1$. Then φ is a finite type invariant of order $(n;k)$.*

2. A_k-equivalence group of spatial graphs.

The following proposition is a natural generalization of [**21**, Lemma] and stems from the fact that a knot with the unknotting number u can be unknotted by changing u crossings of a regular diagram of it [**12**] and [**19**].

Proposition 2.1. *Let $f,g:G\longrightarrow S^3$ be embeddings. If f and g are A_k-equivalent, then g is ambient isotopic to a band sum of f with some links of type k.*

Proof. We consider the embeddings up to ambient isotopy for simplicity. By the assumption there is a finite sequence of embeddings $f=f_0,f_1,\dots,f_n=g$ and orientation preserving embeddings $\varphi_1,\varphi_2,\dots,\varphi_n:B^3\longrightarrow S^3$ such that $(\varphi_i^{-1}(f_{i-1}(G)),\varphi_i^{-1}(f_i(G)))$ is an A_k-move for each i. We shall prove this proposition by induction on n.

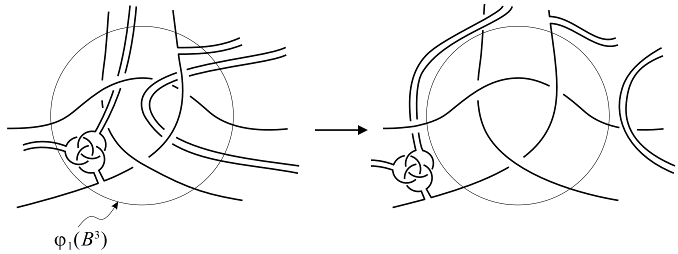


Figure 2.1.

First we consider the case $n = 1$. Let $D_1, D_2, \ldots, D_{k+1}$ be the fixed trivialization of the tangle $\varphi_1^{-1}(f_0(G))$ and $\gamma_j = D_j \cap \partial B^3$ $(j = 1, 2, \ldots, k+1)$. Then $L = \bigcup_j \varphi_1(\gamma_j) \cup (\varphi_1(B^3) \cap f_1(G))$ is a link of type k. By taking a small one-sided collar for each $\varphi_1(\gamma_j)$ in $S^3 - \varphi_1(\mathrm{int}B^3)$, we have mutually disjoint embeddings $b_j : I \times I \longrightarrow S^3$ $(j = 1, 2, \ldots, k+1)$ such that $b_j(I \times I) \cap \varphi_1(B^3) = b_j(I \times \{1\}) = \varphi_1(\gamma_j)$ and $b_j(I \times I) \cap f_0(G) = b_j(I \times I) \cap f_1(G) = b_j(\partial I \times I)$. Then we deform f_0 up to ambient isotopy along the disk $b_j(I \times I) \cup \varphi_1(D_j)$ such that $b_j(I \times I) \cap f_0(G) = b_j(I \times \{0\})$ for each j. Then we have a required band sum $g = F(f_0; \{L\}, \{B\})$, where $B = b_1(I \times I) \cup b_2(I \times I) \cup \cdots \cup b_{k+1}(I \times I)$.

Next suppose that $n > 1$. By the hypothesis of our induction, g is a band sum $F(f_1; \mathcal{L}, \mathcal{B})$, where $\mathcal{L} = \{L_1, L_2, \ldots, L_{n-1}\}$ is a set of links of type k, $\mathcal{B} = \{B_1, B_2, \ldots, B_{n-1}\}$ and each B_i is a union of bands attaching to L_i. Deform $F(f_1; \mathcal{L}, \mathcal{B})$ up to ambient isotopy keeping the image $f_1(G)$ so that neither the associated balls of $\mathcal{L}$ nor the bands in $\mathcal{B}$ intersect $\varphi_1(B^3)$. Note that this deformation is possible, since the tangle $\varphi_1^{-1}(f_1(G))$ is trivial. In fact, sweeping out the associated balls, band-slidings and sweeping out the bands are sufficient. See Figure 2.1. Then by the same arguments as that in the case $n = 1$, we find that f_1 is a band sum $F(f; \{L\}, \{B\})$. Then we have

$$F(F(f; \{L\}, \{B\}); \mathcal{L}, \mathcal{B}) = F(f; \{L\} \cup \mathcal{L}, \{B\} \cup \mathcal{B}).$$

This completes the proof. □

As we mentioned before, the origin of the name 'band trivialization' comes from the following lemma:

Lemma 2.2. *The moves in Figures* 2.2-(i), (ii), (iii) *and* (iv) *are realized by* A_{k+1}*-moves.*

Proof. The move in Figure 2.2-(i) is just a band trivialization of an A_k-move. Hence by the definition it is an A_{k+1}-move. It is easy to see that the moves in Figures 2.2-(ii) and (iii) are generated by the moves in Figure 2.2-(i). To see that the move in Figure 2.2-(iv) is realized by A_{k+1}-moves, we first

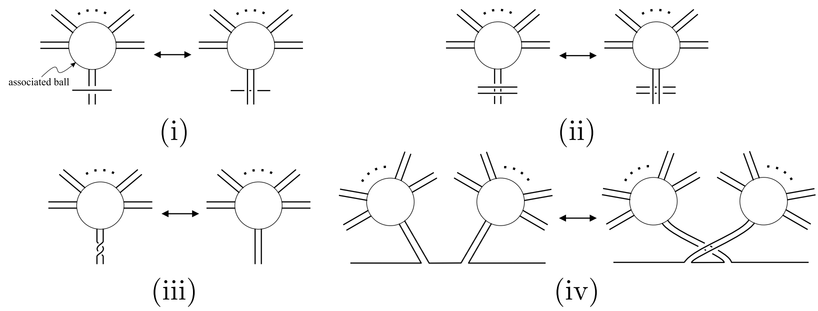


Figure 2.2.

slide the bands as illustrated in Figure 2.3, and then perform the moves in Figure 2.2-(i). □

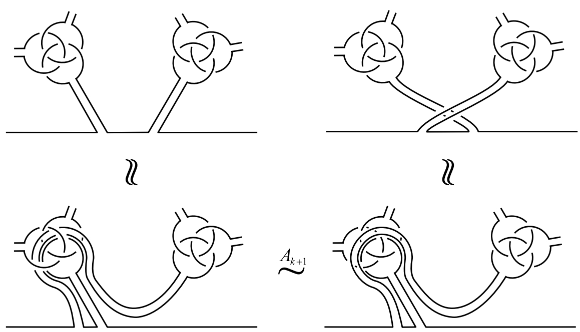


Figure 2.3.

Let $h : G \longrightarrow S^3$ be an embedding and let $[f_1]_k, [f_2]_k \in [h]_{k-1}/(A_k$-equivalence), where $[h]_{k-1}/(A_k$-equivalence) denotes the set of A_k-equivalence classes in $[h]_{k-1}$. Since both f_1 and f_2 are A_{k-1}-equivalent to h, by Proposition 2.1, there are band sums $F(h; \mathcal{L}_i, \mathcal{B}_i) \in [f_i]_k$ of h with links $\mathcal{L}_i$ of type $k-1 (i = 1, 2)$. Suppose that the bands in $\mathcal{B}_1$ and the associated balls of $\mathcal{L}_1$ are disjoint from the bands in $\mathcal{B}_2$ and the associated balls of $\mathcal{L}_2$. Note that up to slight ambient isotopy of $F(h; \mathcal{L}_2, \mathcal{B}_2)$ that preserves $h(G)$ we can always choose the bands and the associated balls so that they satisfy this condition. In the following we assume this condition without explicit mention. Then we have a new band sum $F(h; \mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{B}_1 \cup \mathcal{B}_2)$. We define

$$[f_1]_k +_h [f_2]_k = [F(h; \mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{B}_1 \cup \mathcal{B}_2)]_k.$$

Lemma 2.3. *The sum '$+_h$' above is well-defined.*

Proof. It is sufficient to show for two embeddings $F(h; \mathcal{L}_1, \mathcal{B}_1), F(h; \mathcal{L}'_1, \mathcal{B}'_1) \in [f_1]_k$ that $F(h; \mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{B}_1 \cup \mathcal{B}_2)$ and $F(h; \mathcal{L}'_1 \cup \mathcal{L}_2, \mathcal{B}'_1 \cup \mathcal{B}_2)$ are A_k-equivalent.

Consider a sequence of ambient isotopies and applications of A_k-moves that deforms $F(h;\mathcal{L}_1,\mathcal{B}_1)$ into $F(h;\mathcal{L}_1',\mathcal{B}_1')$. We consider this sequence of deformations together with the links in $\mathcal{L}_2$ and the bands in $\mathcal{B}_2$. Whenever we apply an A_k-move we deform the associated balls of $\mathcal{L}_2$ and the bands in $\mathcal{B}_2$ up to ambient isotopy so that they are away from the 3-ball within which the A_k-move is applied. Thus $F(h;\mathcal{L}_1\cup\mathcal{L}_2,\mathcal{B}_1\cup\mathcal{B}_2)=F(F(h;\mathcal{L}_1,\mathcal{B}_1);\mathcal{L}_2,\mathcal{B}_2)$ is A_k-equivalent to a band sum $F(F(h;\mathcal{L}_1',\mathcal{B}_1');\mathcal{L}_2',\mathcal{B}_2')$ for some $\mathcal{L}_2'$ and $\mathcal{B}_2'$. Compare the band sums $F(F(h;\mathcal{L}_1',\mathcal{B}_1');\mathcal{L}_2',\mathcal{B}_2')$ and $F(h;\mathcal{L}_1'\cup\mathcal{L}_2,\mathcal{B}_1'\cup\mathcal{B}_2)=F(F(h;\mathcal{L}_1',\mathcal{B}_1');\mathcal{L}_2,\mathcal{B}_2)$. We have that the links in $\mathcal{L}_2'$ are ambient isotopic to the links in $\mathcal{L}_2$. It follows from Lemma 2.2 that the bands in $\mathcal{B}_2'$ can be deformed into the position of the bands in $\mathcal{B}_2$ by band slidings and A_k-moves. Thus these two are A_k-equivalent. □

Theorem 2.4. *The set $[h]_{k-1}/(A_k$-equivalence$)$ forms an abelian group under '$+_h$' with the unit element $[h]_k$.*

We denote this group by $\mathcal{G}_k(h;G)$ and call it the *A_k-equivalence group* of the spatial embeddings of G with the unit element $[h]_k$.

Remark. Note that for any graph G and any embedding $h:G\longrightarrow S^3$, $[h]_1$ is equal to the set of all embeddings of G into S^3. In [**20**], the second author called $\mathcal{G}_2(h;G)$ a *graph homology group* and gave a practical method of calculating this group.

Proof. We consider embeddings up to ambient isotopy for simplicity. It is sufficient to show that for any $[f]_k\in[h]_{k-1}/(A_k$-equivalence), there is an inverse of $[f]_k$. Since f and h are A_{k-1}-equivalent, by Proposition 2.1, f and h are band sums $F(h;\mathcal{L},\mathcal{B})$ and $F(f;\mathcal{L}',\mathcal{B}')$ respectively, where $\mathcal{L}$ and $\mathcal{L}'$ are sets of links of type $k-1$. Thus we have $h=F(F(h;\mathcal{L},\mathcal{B});\mathcal{L}',\mathcal{B}')$. Then, by using Lemma 2.2, we deform the associated balls of $\mathcal{L}'$ and the bands in $\mathcal{B}'$ up to A_k-equivalence so that they are disjoint form the associated balls of $\mathcal{L}$ and the bands in $\mathcal{B}$. Thus we see that $h=F(F(h;\mathcal{L},\mathcal{B});\mathcal{L}',\mathcal{B}')$ is A_k-equivalent to a band sum $F(h;\mathcal{L}\cup\mathcal{L}'',\mathcal{B}\cup\mathcal{B}'')$ for some $\mathcal{L}''$ and $\mathcal{B}''$ (for example see Figure 2.4). Thus we have

$$\begin{aligned}[f]_k+_h[F(h;\mathcal{L}'',\mathcal{B}'')]_k&=[F(h;\mathcal{L},\mathcal{B})]_k+_h[F(h;\mathcal{L}'',\mathcal{B}'')]_k\\&=[F(h;\mathcal{L}\cup\mathcal{L}'',\mathcal{B}\cup\mathcal{B}'')]_k\\&=[h]_k.\end{aligned}$$

This implies that $[F(h;\mathcal{L}'',\mathcal{B}'')]_k$ is an inverse of $[f]_k$. □

Theorem 2.5. *Let $h_1,h_2:G\longrightarrow S^3$ be A_{k-1}-equivalent embeddings. Then the groups $\mathcal{G}_k(h_1;G)$ and $\mathcal{G}_k(h_2;G)$ are isomorphic.*

Proof. We define a map $\phi:\mathcal{G}_k(h_1;G)\longrightarrow\mathcal{G}_k(h_2;G)$ by $\phi([f]_k)=[f]_k-_{h_2}[h_1]_k$, where $[x]_k-_{h_2}[y]_k$ denotes $[x]_k+_{h_2}(-[y]_k)$. Clearly this map is a bijection. We shall prove that ϕ is a homomorphism. Let $[f_i]_k\in\mathcal{G}_k(h_1;G)$ $(i=$

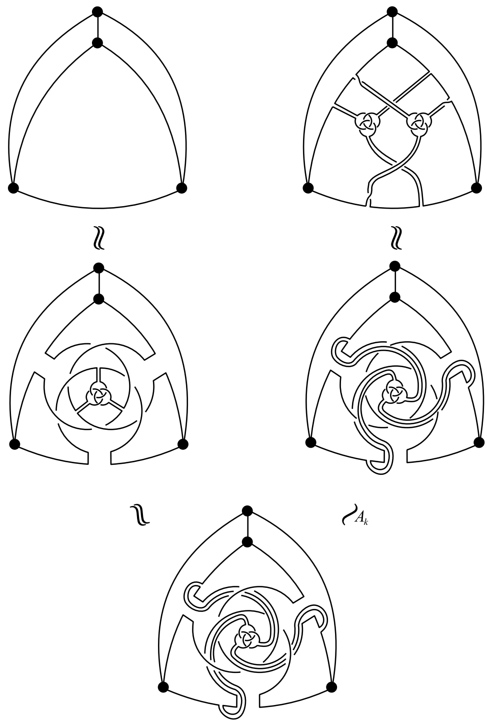


Figure 2.4.

$1,2)$. Then $f_i = F(h_1; \mathcal{L}_i, \mathcal{B}_i)$ where $\mathcal{L}_i$ is a set of links of type $k-1$ $(i = 1,2)$. Since h_1 and h_2 are A_{k-1}-equivalent we see that $h_1 = F(h_2; \mathcal{L}, \mathcal{B})$ where $\mathcal{L}$ is a set of links of type $k-1$. Thus we have $f_i = F(F(h_2; \mathcal{L}, \mathcal{B}); \mathcal{L}_i, \mathcal{B}_i)$ $(i = 1,2)$. By using Lemma 2.2, we deform f_i up to A_k-equivalence so that the associated balls of $\mathcal{L}_i$ and the bands in $\mathcal{B}_i$ are disjoint from the associated balls of $\mathcal{L}$ and the bands in $\mathcal{B}$ for $i = 1,2$. We may further assume that the associated balls of $\mathcal{L}_1$ and the bands in $\mathcal{B}_1$ are disjoint from the associated balls of $\mathcal{L}_2$ and the bands in $\mathcal{B}_2$. Then we have

$$\begin{aligned}
\phi([f_1]_k +_{h_1} [f_2]_k) &= \phi([F(F(h_2; \mathcal{L}, \mathcal{B}); \mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{B}_1 \cup \mathcal{B}_2)]_k) \\
&= [F(F(h_2; \mathcal{L}, \mathcal{B}); \mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{B}_1 \cup \mathcal{B}_2)]_k -_{h_2} [h_1]_k \\
&= [F(h_2; \mathcal{L} \cup \mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{B} \cup \mathcal{B}_1 \cup \mathcal{B}_2)]_k -_{h_2} [F(h_2; \mathcal{L}, \mathcal{B})]_k \\
&= [F(h_2; \mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{B}_1 \cup \mathcal{B}_2)]_k,
\end{aligned}$$

and for each i $(i = 1,2)$,

$$\begin{aligned}
\phi([f_i]_k) &= \phi([F(F(h_2; \mathcal{L}, \mathcal{B}); \mathcal{L}_i, \mathcal{B}_i)]_k) \\
&= [F(F(h_2; \mathcal{L}, \mathcal{B}); \mathcal{L}_i, \mathcal{B}_i)]_k -_{h_2} [h_1]_k \\
&= [F(h_2; \mathcal{L} \cup \mathcal{L}_i, \mathcal{B} \cup \mathcal{B}_i)]_k -_{h_2} [F(h_2; \mathcal{L}, \mathcal{B})]_k \\
&= [F(h_2; \mathcal{L}_i, \mathcal{B}_i)]_k.
\end{aligned}$$

Thus we have $\phi([f_1]_k +_{h_1} [f_2]_k) = \phi([f_1]_k) +_{h_2} \phi([f_2]_k)$. □

Proposition 2.6. *The projection* $p : [h]_{k-1} \longrightarrow [h]_{k-1}/(A_k\text{-}equivalence) = \mathcal{G}_k(h;G)$ *is a finite type* A_k*-equivalence invariant of order* $(1;k)$.

Proof. It is clear that p is an A_k-equivalence invariant. We shall prove that p is finite type of order $(1;k)$. Let $f \in [h]_{k-1}$ be an embedding and $F(f;\{L_1,L_2\},\{B_1,B_2\})$ a band sum of f with links L_1, L_2 of type $k-1$. Then it is sufficient to show that

$$\sum_{X\subset\{1,2\}} (-1)^{|X|} p\left(F\left(f;\bigcup_{i\in X}\{L_i\},\bigcup_{i\in X}\{B_i\}\right)\right) = [h]_k.$$

Let $\phi : \mathcal{G}_k(f;G) \longrightarrow \mathcal{G}_k(h;G)$ be the isomorphism defined by $\phi([g]_k) = [g]_k -_h [f]_k$. Then we have

$$\begin{aligned}
&\phi\Big([F(f;\emptyset,\emptyset)]_k -_f [F(f;\{L_1\},\{B_1\})]_k \\
&\qquad -_f [F(f;\{L_2\},\{B_2\})]_k +_f [F(f;\{L_1,L_2\},\{B_1,B_2\})]_k\Big) \\
&= ([F(f;\emptyset,\emptyset)]_k -_h [f]_k) -_h ([F(f;\{L_1\},\{B_1\})]_k -_h [f]_k) \\
&\quad -_h ([F(f;\{L_2\},\{B_2\})]_k -_h [f]_k) +_h ([F(f;\{L_1,L_2\},\{B_1,B_2\})]_k -_h [f]_k) \\
&= [F(f;\emptyset,\emptyset)]_k -_h [F(f;\{L_1\},\{B_1\})]_k -_h [F(f;\{L_2\},\{B_2\})]_k \\
&\quad +_h [F(f;\{L_1,L_2\},\{B_1,B_2\})]_k \\
&= \sum_{X\subset\{1,2\}} (-1)^{|X|} p\left(F\left(f;\bigcup_{i\in X}\{L_i\},\bigcup_{i\in X}\{B_i\}\right)\right).
\end{aligned}$$

Since

$$[F(f;\{L_1,L_2\},\{B_1,B_2\})]_k = [F(f;\{L_1\},\{B_1\})]_k +_f [F(f;\{L_2\},\{B_2\})]_k,$$

we have

$$\begin{aligned}
&\phi\Big([F(f;\emptyset,\emptyset)]_k -_f [F(f;\{L_1\},\{B_1\})]_k -_f [F(f;\{L_2\},\{B_2\})]_k \\
&\qquad +_f [F(f;\{L_1,L_2\},\{B_1,B_2\})]_k\Big) \\
&= \phi([f]_k) \\
&= [h]_k.
\end{aligned}$$

This completes the proof. □

Proof of Theorem 1.1. The ‘only if’ part is clear. We show the ‘if’ part. Let f and g be embeddings in $[h]_{k-1}$. Suppose that any finite type invariant of order $(1;k)$ takes the same value on f and g. Then by Proposition 2.6 we have $p(f) = p(g)$, where $p : [h]_{k-1} \longrightarrow [h]_{k-1}/(A_k\text{-equivalence}) = \mathcal{G}_k(h;G)$ is the projection. Hence we have $[f]_k = [g]_k$. This completes the proof. □

3. A_k-equivalence group of knots.

In this section we only consider the case that the graph G is homeomorphic to a disjoint union of circles. Let $G = S^1_1 \cup S^1_2 \cup \cdots \cup S^1_\mu$. Then there is a natural correspondence between the ambient isotopy classes of the embeddings of G into S^3 and the ambient isotopy classes of the ordered oriented μ-component links in S^3. Therefore instead of specifying an embedding $h : S^1_1 \cup S^1_2 \cup \cdots \cup S^1_\mu \longrightarrow S^3$, we denote by L the image $h(S^1_1 \cup S^1_2 \cup \cdots \cup S^1_\mu)$ and consider it together with the orientation of each component and the ordering of the components. Thus $\mathcal{G}_k(L)$ denotes the A_k-equivalence group $\mathcal{G}_k(h; S^1_1 \cup S^1_2 \cup \cdots \cup S^1_\mu)$ with the unit element $[h]_k$.

Theorem 3.1. *Let O be a trivial knot. Then for any oriented knot K, $\mathcal{G}_k(O)$ and $\mathcal{G}_k(K)$ are isomorphic.*

Remark. For a graph $G(\neq S^1)$ and embeddings $h, h' : G \longrightarrow S^3$, $\mathcal{G}_k(h; G)$ and $\mathcal{G}_k(h'; G)$ are not always isomorphic. In fact there are two-component links L_1 and L_2 such that $\mathcal{G}_3(L_1) \cong \mathbb{Z} \oplus \mathbb{Z}$ and $\mathcal{G}_3(L_2) \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}_2$ [**16**].

Proof. We define a map $\phi : \mathcal{G}_k(O) \longrightarrow \mathcal{G}_k(K)$ by

$$\phi([F(O; \mathcal{L}, \mathcal{B})]_k) = [K \# F(O; \mathcal{L}, \mathcal{B})]_k$$

for each $[F(O; \mathcal{L}, \mathcal{B})]_k \in \mathcal{G}_k(O)$, where $\mathcal{L}$ is a set of links of type $k-1$ and $\#$ means the connected sum of oriented knots. Clearly this is well-defined. By Lemma 2.2, any band sum $F(K; \mathcal{L}, \mathcal{B})$ of K with links $\mathcal{L}$ of type $k-1$ is A_k-equivalent to $K \# F(O; \mathcal{L}', \mathcal{B}')$ for some links $\mathcal{L}'$ of type $k-1$ and $\mathcal{B}'$. Hence ϕ is surjective. For $[F(O; \mathcal{L}_i, \mathcal{B}_i)]_k \in \mathcal{G}_k(O)$ $(i = 1, 2)$, we have

$$\begin{aligned}
&\phi([F(O; \mathcal{L}_1, \mathcal{B}_1)]_k +_O [F(O; \mathcal{L}_2, \mathcal{B}_2)]_k) \\
&= \phi([F(O; \mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{B}_1 \cup \mathcal{B}_2)]_k) \\
&= [K \# F(O; \mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{B}_1 \cup \mathcal{B}_2)]_k \\
&= [F(K; \mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{B}_1 \cup \mathcal{B}_2)]_k \\
&= [F(K; \mathcal{L}_1, \mathcal{B}_1)]_k +_K [F(K; \mathcal{L}_2, \mathcal{B}_2)]_k \\
&= [K \# F(O; \mathcal{L}_1, \mathcal{B}_1)]_k +_K [K \# F(O; \mathcal{L}_2, \mathcal{B}_2)]_k \\
&= \phi([F(O; \mathcal{L}_1, \mathcal{B}_1)]_k) +_K \phi([F(O; \mathcal{L}_2, \mathcal{B}_2)]_k).
\end{aligned}$$

This implies that ϕ is a homomorphism. In order to complete the proof, we show that ϕ is injective. Suppose that $[K \# F(O; \mathcal{L}, \mathcal{B})]_k = [K]_k$. By Lemma 3.2, there is a knot K' such that $[K' \# K]_k = [O]_k$. Then we have

$$\begin{aligned}
[F(O; \mathcal{L}, \mathcal{B})]_k &= [(K' \# K) \# F(O; \mathcal{L}, \mathcal{B})]_k \\
&= [K' \# (K \# F(O; \mathcal{L}, \mathcal{B}))]_k \\
&= [K' \# K]_k \\
&= [O]_k.
\end{aligned}$$

This implies that $\ker \phi = \{[O]_k\}$. $\square$

Habiro originated 'clasper theory' and showed Lemma 3.2 for C_k-moves [**3**] and [**4**]. The following proof is a translation of his proof in terms of band sum description of knots:

Lemma 3.2. *For any knot K and any integer $k \geq 1$, there is a knot K' such that $K'\#K$ is A_k-equivalent to a trivial knot.*

Proof. We shall prove this by induction on k. The case $k = 1$ is clear. Suppose that there is a knot K' such that $K'\#K$ is A_{k-1}-equivalent to a trivial knot O ($k > 1$). By Proposition 2.1, we may assume that $O = F(K'\#K; \mathcal{L}, \mathcal{B})$, where $\mathcal{L}$ is a set of links of type $k-1$. Then, by Lemma 2.2, we see that $F(K'\#K; \mathcal{L}, \mathcal{B})$ is A_k-equivalent to some $K\#F(K'; \mathcal{L}, \mathcal{B}')$. This completes the proof. $\square$

Let $\mathcal{K}_k$ be the set of A_k-equivalence classes of all oriented knots. For $[K]_k, [K']_k \in \mathcal{K}_k$, we define $[K]_k + [K']_k = [K\#K']_k$. Then the following, shown by Habiro [**3**],[**4**] in the case that A_k-moves coincide with C_k-moves, is an immediate consequence of Lemma 3.2.

Theorem 3.3. *The set $\mathcal{K}_k$ forms an abelian group under '$+$' with the unit element $[O]_k$, where O is a trivial knot.*

4. Generalized A_k-move.

In this section, we define a generalized A_k-move. For this move, several results similar to that in Sections 1, 2 and 3 hold.

Let T and S be trivial tangles such that (T, S) and (S, T) are equivalent. Let $t_1, t_2, \ldots, t_n$ and $s_1, s_2, \ldots, s_n$ be the components of T and S respectively. An $A_1(T, S)$-*move* is this local move (T, S). Suppose that $A_k(T, S)$-moves are defined. For each $A_k(T, S)$-move (T_k, S_k) we choose a trivialization of T_k and fix it. Then the band trivializations of (T_k, S_k) with respect to the trivialization are called $A_{k+1}(T, S)$-*moves.* Let (T_k, S_k) be an $A_k(T, S)$-move and $D_1, D_2, \ldots, D_{n+k-1}$ the fixed trivialization of $T_k = t_1 \cup t_2 \cup \cdots \cup t_{n+k-1}$. We set $\alpha = \partial B^3 \cap (D_1 \cup D_2 \cup \cdots \cup D_{n+k-1})$ and $\beta = S_k$. A link L in S^3 is called *type* $(k; (T, S))$ if there is an orientation preserving embedding $\varphi : B^3 \longrightarrow S^3$ such that $L = \varphi(\alpha \cup \beta)$. Then the pair (α, β) is called a *link model of* L. As in Section 1, $A_k(T, S)$-move gives an equivalence relation, $A_k(T, S)$-*equivalence*, on the set of all embeddings of G into S^3. For an embedding $f : G \longrightarrow S^3$, let $[f]_k$ denote the $A_k(S, T)$-equivalence class of f. Let $h : G \longrightarrow S^3$ be an embedding and H an abelian group. Let $\varphi : [h]_{k-1} \longrightarrow H$ be an invariant. We can define that φ is a *finite type invariant of order* $(n; k; (T, S))$ as in Section 1.

By the arguments similar to that in Sections 1, 2 and 3, we have the following five theorems:

Theorem 4.1. *Let $f, g : G \longrightarrow S^3$ be $A_{k-1}(T, S)$-equivalent embeddings. Then they are $A_k(T, S)$-equivalent if and only if $\varphi(f) = \varphi(g)$ for any finite type $A_k(T, S)$-equivalence invariant φ of order $(1; k; (T, S))$.*

Let $h : G \longrightarrow S^3$ be an embedding. For $[f_1]_k, [f_2]_k \in [h]_{k-1}/(A_k(T, S)$-equivalence), we can define $[f_1]_k +_h [f_2]_k$ as in Section 2, and we have:

Theorem 4.2. *The set $[h]_{k-1}/(A_k(T, S)$-equivalence) forms an abelian group under '$+_h$' with the unit element $[h]_k$.*

We denote this group by $\mathcal{G}_{k(T,S)}(h; G)$ and call it the $A_k(T, S)$-*equivalence group* of the spatial embeddings of G with the unit element $[h]_k$.

Theorem 4.3. *Let $h_1, h_2 : G \longrightarrow S^3$ be $A_{k-1}(T, S)$-equivalent embeddings. Then the groups $\mathcal{G}_{k(T,S)}(h_1; G)$ and $\mathcal{G}_{k(T,S)}(h_2; G)$ are isomorphic.*

For an embedding $h : S^1 \longrightarrow S^3$, let $K = h(S^1)$ and let $\mathcal{G}_{k(T,S)}(K)$ denote the $A_k(T, S)$-equivalence group $\mathcal{G}_{k(T,S)}(h; S^1)$ with the unit element $[h]_k$.

Theorem 4.4. *Let O be a trivial knot. If any two knots are $A_1(T, S)$-equivalent, then for any oriented knot K, $\mathcal{G}_{k(T,S)}(O)$ and $\mathcal{G}_{k(T,S)}(K)$ are isomorphic.*

Let $\mathcal{K}_{k(T,S)}$ be the set of $A_k(T, S)$-equivalence classes of all oriented knots. For $[K]_k, [K']_k \in \mathcal{K}_{k(T,S)}$, we define $[K]_k + [K']_k = [K \# K']_k$.

Theorem 4.5. *If any two knots are $A_1(T, S)$-equivalent, then the set $\mathcal{K}_{k(T,S)}$ forms an abelian group under '$+$' with the unit element $[O]_k$, where O is a trivial knot.*

Remark. If (T, S) is the #-move defined by Murakami [**6**], then $\mathcal{K}_{k(T,S)}$ is an abelian group.

5. A_k-moves and Vassiliev invariants.

Let G be a finite graph. We give and fix orientations of the edges of G. Let $\mathcal{E}$ be the set of the ambient isotopy classes of the embeddings of G into S^3. Let $\mathbb{Z}\mathcal{E}$ be the free abelian group generated by the elements of $\mathcal{E}$. A *crossing vertex* is a double point of a map from G to S^3 as in Figure 5.1. An *i-singular embedding* is a map from G to S^3 whose multiple points are exactly i crossing vertices. By the formula in Figure 5.2 we identify an i-singular embedding with an element in $\mathbb{Z}\mathcal{E}$. Let $\mathcal{R}_i$ be the subgroup of $\mathbb{Z}\mathcal{E}$ generated by all i-singular embeddings. Note that $\mathcal{R}_i$ is independent of the choices of the edge orientations. Let H be an abelian group. Let $\varphi : \mathcal{E} \longrightarrow H$ be a map. Let $\widetilde{\varphi} : \mathbb{Z}\mathcal{E} \longrightarrow H$ be the natural extension of φ. We say that φ is a *Vassiliev invariant of order n* if $\widetilde{\varphi}(\mathcal{R}_{n+1}) = \{0\}$. Let $\iota : \mathcal{E} \longrightarrow \mathbb{Z}\mathcal{E}$ be the natural inclusion map and $\pi_i : \mathbb{Z}\mathcal{E} \longrightarrow \mathbb{Z}\mathcal{E}/\mathcal{R}_i$ the quotient homomorphism.

Let $u_{i-1} = \pi_i \circ \iota : \mathcal{E} \longrightarrow \mathbb{Z}\mathcal{E}/\mathcal{R}_i$ be the composition map. Then φ is a Vassiliev invariant of order n if and only if there is a homomorphism $\hat{\varphi} : \mathbb{Z}\mathcal{E}/\mathcal{R}_{n+1} \longrightarrow H$ such that $\varphi = \hat{\varphi} \circ u_n$. In the following we sometimes do not distinguish between an embedding and its ambient isotopy class so long as no confusion occurs.

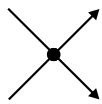


Figure 5.1.

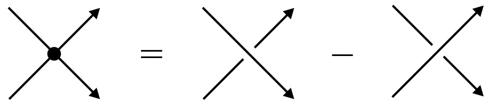


Figure 5.2.

Theorem 5.1. *Let $f, g : G \longrightarrow S^3$ be A_{k+1}-equivalent embeddings. Then $u_k(f) = u_k(g)$.*

By using induction on k, we see that an A_k-move (T, S) is a $(k+1)$-component Brunnian local move, i.e., $T - t$ and $S - s$ are ambient isotopic in B^3 relative ∂B^3 for any $t \in T$ and $s \in S$ with $\partial t = \partial s$ [**15**]. It is not hard to see that if two embeddings f and g are related by a $(k+1)$-component Brunnian local move, then f and g are *k-similar*, where k-similar is an equivalence relation defined by the first author [**13**]. Therefore, we note that Theorem 5.1 follows from [**1**] or [**9**]. However we give a self-contained proof here.

Let T be a tangle. Let $\mathcal{H}(T)$ be the set of all (possibly nontrivial) tangles that are homotopic to T relative to ∂B^3. Let $\mathcal{E}(T)$ be the quotient of $\mathcal{H}(T)$ by the ambient isotopy relative to ∂B^3. Then $\mathbb{Z}\mathcal{E}(T)$, i-singular tangles and $\mathcal{R}_i(T) \subset \mathbb{Z}\mathcal{E}(T)$ are defined as above.

Proof. It is sufficient to show for each A_k-move (T, S) that $T - S$ is an element in $\mathcal{R}_k(T)$. We show this by induction on k. The case $k = 1$ is clear. Recall that an A_k-move (T, S) is a band trivialization of an A_{k-1}-move, say (T', S'). Then we have that $T - S = X_1 - X_2$ where X_1 and X_2 are 1-singular tangles in Figure 5.3. Let Y_1 and Y_2 be 1-singular tangles in Figure 5.4. It is clear that $Y_1 - Y_2 = 0$. Thus we have $T - S = (X_1 - Y_1) - (X_2 - Y_2)$. By the induction hypothesis we see that $S' - T'$ is an element of $\mathcal{R}_{k-1}(S')$. Therefore both $X_1 - Y_1$ and $X_2 - Y_2$ are elements of $\mathcal{R}_k(T)$. $\square$

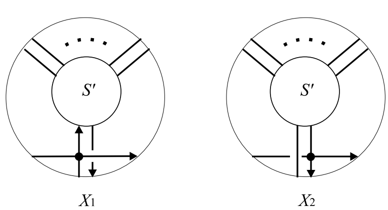


Figure 5.3.

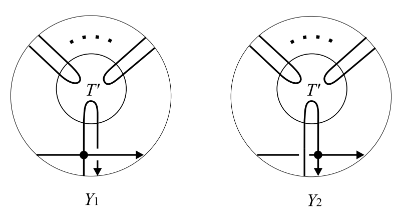


Figure 5.4.

Proof of Proposition 1.3. It is sufficient to show that

$$\sum_{X\subset\{1,2,\ldots,n+1\}}(-1)^{|X|}F\left(f;\bigcup_{i\in X}\{L_i\},\bigcup_{i\in X}\{B_i\}\right)$$

is an element of $\mathcal{R}_{(n+1)(k-1)}$. We show this together with some additional claims by induction on n. First consider the case $n=0$. Then we have by Theorem 5.1 and its proof that $F(f;\emptyset,\emptyset)-F(f;\{L_1\},\{B_1\})$ is a sum of $(k-1)$-singular embeddings each of which has all crossing vertices in the associated ball. Note that these $(k-1)$-singular embeddings are natural extensions of the $(k-1)$-singular tangles that express the difference of the A_{k-1}-move, and these $(k-1)$-singular tangles depends only on the link L_1. Next we consider the general case. Note that

$$\begin{aligned}
&\sum_{X\subset\{1,2,\ldots,n+1\}}(-1)^{|X|}F\left(f;\bigcup_{i\in X}\{L_i\},\bigcup_{i\in X}\{B_i\}\right)\\
&=\sum_{X\subset\{1,2,\ldots,n\}}(-1)^{|X|}F\left(f;\bigcup_{i\in X}\{L_i\},\bigcup_{i\in X}\{B_i\}\right)\\
&\quad-\sum_{X\subset\{1,2,\ldots,n\}}(-1)^{|X|}F\left(f;\bigcup_{i\in X}\{L_i\}\cup\{L_{n+1}\},\bigcup_{i\in X}\{B_i\}\cup\{B_{n+1}\}\right).
\end{aligned}$$

By the hypothesis we have both

$$\sum_{X\subset\{1,2,\ldots,n\}}(-1)^{|X|}F\left(f;\bigcup_{i\in X}\{L_i\},\bigcup_{i\in X}\{B_i\}\right)$$

and

$$\sum_{X\subset\{1,2,\ldots,n\}}(-1)^{|X|}F\left(f;\bigcup_{i\in X}\{L_i\}\cup\{L_{n+1}\},\bigcup_{i\in X}\{B_i\}\cup\{B_{n+1}\}\right)$$

are sums of $n(k-1)$-singular embeddings and they differ only by the band sum of L_{n+1}. Therefore we have

$$\sum_{X\subset\{1,2,\ldots,n+1\}}(-1)^{|X|}F\left(f;\bigcup_{i\in X}\{L_i\},\bigcup_{i\in X}\{B_i\}\right)$$

is a sum of $(n(k-1)+(k-1))$-singular embeddings. □

References

[1] M. Gusarov, *On n-equivalence of knots and invariants of finite degree*, in 'Topology of Manifolds and Varieties' (O. Viro, ed.), American Mathematical Society, Providence, RI, 173-192, 1994, MR 96i:57005, Zbl 0865.57007.

[2] K. Habiro, *Aru musubime no kyokusyo henkei no zoku ni tuite* (Japanese), Master's Thesis, University of Tokyo, 1994.

[3] ______, *Claspers and the Vassiliev skein modules*, Ph.D. Thesis, University of Tokyo, 1997.

[4] ______, *Claspers and finite type invariants of links*, Geom. Topol., **4** (2000), 1-83, MR 2001g:57020, Zbl 0941.57015.

[5] T. Motohashi and K. Taniyama, *Delta unknotting operation and vertex homotopy of spatial graphs*, in 'Proceedings of Knots '96 Tokyo' (S. Suzuki, ed.), 185-200, World Scientific Publ. Co., 1997, MR 99i:57021, Zbl 0960.57003.

[6] H. Murakami, *Some metrics of classical knots*, Math. Ann., **270** (1985), 35-45, MR 86g:57007, Zbl 0535.57005.

[7] ______, *Vassiliev invariants of type two for a link*, Proc. Amer. Math. Soc., **124** (1996), 3889-3896, MR 97b:57007, Zbl 0874.57004.

[8] H. Murakami and Y. Nakanishi, *On a certain move generating link-homology*, Math. Ann., **284** (1989), 75-89, MR 90f:57007, Zbl 0646.57005.

[9] Y. Ohyama, *Vassiliev invariants and similarity of knots*, Proc. Amer. Math. Soc., **123** (1995), 287-291, MR 95f:57018, Zbl 0847.57012.

[10] Y. Ohyama, K. Taniyama and S. Yamada, *Vassiliev invariants of unknotting number one knots*, Tokyo J. Math., **25**(**1**) (2002), 17-31, MR 2003d:57032, Zbl 1002.57030.

[11] T. Stanford, *Finite-type invariants of knots, links and graphs*, Topology, **35** (1996), 1027-1050, MR 97i:57009, Zbl 0863.57005.

[12] S. Suzuki, *Local knots of* 2*-spheres in* 4*-manifolds*, Proc. Japan Acad., **45** (1969), 34-38, MR 40 #2101, Zbl 0176.53601.

[13] K. Taniyama, *On similarity of links*, Gakujutu Kenku (issued by School of Education, Waseda University), **14** (1992), 33-36.

[14] ______, *Homology classification of spatial embeddings of a graph*, Topology Appl., **65** (1995), 205-228, MR 97c:57003, Zbl 0843.57012.

[15] K. Taniyama and A. Yasuhara, *Realization of knots and links in a spatial graph*, Topology Appl., **112** (2001), 87-109, MR 2002e:57005, Zbl 0968.57001.

[16] ______, *Clasp-pass moves on spatial graphs*, Topology Appl., **122** (2002), 501-529, CMP 1 911 697, Zbl 1001.57011.

[17] ______, *Band description of knots and Vassiliev invariants*, Math. Proc. Cambridge Philos. Soc., **133** (2002), 325-343, CMP 1 912 405.

[18] W.T. Wu, *On the isotopy of a complexes in a Euclidean space*, I, Sci. Sin., **9** (1960), 21-46, MR 23 #A1381, Zbl 0121.18102.

[19] M. Yamamoto, *Knots in spatial embeddings of the complete graph on four vertices*, Topology Appl., **36** (1990), 291-298, MR 91m:57007, Zbl 0721.57005.

[20] A. Yasuhara, *Disk/band surface and spatial-graph homology*, Topology Appl., **69** (1996), 173-191, MR 97b:57003, Zbl 0860.57004.

[21] ———, *Delta-unknotting operation and adaptability of certain graphs*, in 'Proceedings of Knots '96 Tokyo' (S. Suzuki, ed.), 115-121, World Scientific Publ. Co., 1997, MR 99j:57004, Zbl 0960.57004.

Received July 11, 2000 and revised July 30, 2001.

Department of Mathematics
Waseda Universitry
Shinjuku-ku
Tokyo 169-8050
Japan
E-mail address: taniyama@waseda.jp

Department of Mathematics
Tokyo Gakugei University
Koganei
Tokyo 184-8501
Japan
E-mail address: yasuhara@u-gakugei.ac.jp

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 BibTeX 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 will 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.

PACIFIC JOURNAL OF MATHEMATICS

Volume 211 No. 1 September 2003

0030-8730(200309)211:1;1-N