BRAUER-TYPE RESULTS ON SEMIGROUPS OVER $p$-ADIC FIELDS

Janez Bernik

In this paper we show that every central simple algebra $A$ over $\mathbb{Q}_p$, generated by a multiplicative semigroup $S \subset A$ with the property that the minimal polynomial of every element in $S$ splits over $\mathbb{Q}_p$, is isomorphic to $M_n(\mathbb{Q}_p)$. If, in addition, $S \subset A^*$ is a compact group, then it contains a commutative normal subgroup of finite index.

1. Introduction.

In this paper we consider the following problem. Let $k < K$ be fields. Suppose $S \subset M_n(K)$ is an absolutely irreducible multiplicative semigroup with the property that the spectrum of every $a \in S$ is contained in $k$. Does it follow that $S$ is simultaneously realizable in $M_n(k)$, that is, does there exist a $p \in GL_n(K)$ such that $pSp^{-1} \subset M_n(k)$? The overfield $K$ plays no essential role in this situation. Consider the $k$-algebra $A$ generated by $S$. It is easy to see that it is a central simple algebra over $k$ (see [7]). If one can show that $A$ is isomorphic to $M_n(k)$, then this isomorphism can be extended to an inner isomorphism of $M_n(K)$ by Skolem-Noether theorem. So we consider the more intrinsic version of the question above. Suppose $A$ is a central simple algebra over $k$ and $S \subset A$ a multiplicative semigroup that generates $A$ as a $k$-vector space and has the property that the minimal polynomial of every element in $S$ splits over $k$. Does it follow $A \simeq M_n(k)$? If we exclude the case when the Brauer group of $k$ is trivial, then the question has no apparent answer. If $S$ is a finite subgroup of $A^*$ and char $(k) = 0$, then the answer is affirmative by Brauer’s theorem on splitting fields (see [3, Thm. 41.1]).

The general semigroup case was considered for some particular fields. In [7] it is shown that the answer is affirmative in the special case $k = \mathbb{R}$ with no additional assumptions on $S$ (see also [9] for some related results). In this paper we consider the case $k = \mathbb{Q}_p$ for every rational prime number $p$.

The proof proceeds in two steps. First we consider the special case when $S$ is a compact subgroup of $A^*$. The crucial part in this case is the fact that the Lie algebra of $S$ is commutative. This however is not true if $k$ is a general $p$-adic field (see the example at the end of the paper). The second step is to reduce the problem from arbitrary semigroup with the desired property to
the compact group case. Although we state this result for $k = \mathbb{Q}_p$ only, the reader can easily verify that this reduction works for any $p$-adic field.

2. The results.

We start with a simple observation that holds true for any field $k$. Let $\mu_n(k)$ denote the group of $n$-th roots of unity in $k$.

**Proposition 1.** Let $A$ be a central simple algebra over a field $k$. Suppose $A$ is spanned over $k$ by a center-by-finite group $S \subset A^*$ with the property that the minimal polynomial of every $s \in S$ splits over $k$. Then $A \simeq M_n(k)$.

**Proof.** We assume with no loss of generality that $k^* < S$ and let $r$ denote the exponent of the finite group $S/k^*$. The conditions of the proposition imply that every element $a \in S$ is a scalar multiple of an element $b \in S$ where $b^r = 1$. Now, let $a, b \in S$ be two elements of order dividing $r$. Their product $c = ab$ may be of order greater than $r$ but $c = \alpha d$, $\alpha \in k^*$, $d \in S$, $d^r = 1$. Applying the reduced norm we see that

$$\alpha^{nr} = nr(c^r) = nr((ab)^r) = nr(a^r)nr(b^r) = 1$$

where $n$ is the reduced degree of $A$ over $k$. Thus we have shown that $S_1 = \{ab; \alpha \in \mu_{nr}(k), b \in S, b^r = 1\}$ is a subgroup of $S$ of bounded period which clearly spans $A$. By a well-known result by Burnside this implies $S_1$ is finite (the reduced trace on $A$ is nondegenerate and it takes only finitely many values on $S_1$).

Suppose now $\text{char}(k) = 0$. If $m$ is the exponent of $S_1$, then it is easy to see that $k$ contains a primitive $m$-th root of unity so we can apply the Brauer’s theorem on splitting fields and the proof is complete in this case. If $\text{char}(k) \neq 0$, then the claim follows from ([4], Proof of Corollary 7.11, p. 148).

It should be mentioned that there is an alternative but less elementary way to prove this result by using Schur’s theory of projective representations of finite groups. Observe also that in the particular case when $\text{char}(k) = 0$ and the only roots of unity in $k$ are $\pm 1$ (e.g., $k = \mathbb{R}, \mathbb{Q}_2, \mathbb{Q}_3$) $S_1$ is of exponent 2, therefore commutative and $A \simeq k$.

In what follows we will need a slight generalization of the previous proposition. We shall also use repeatedly the following fact. If one wants to show the triviality of a central simple algebra $A$ in the Brauer group of $k$, then it suffices to consider the central simple algebra $eAe$ where $e \in A$ ia a nonzero idempotent. The proof may be found in [5].

**Corollary 2.** Let $A$ be a central simple algebra over a field $k$. Suppose $A$ is spanned over $k$ by a group $S \subset A^*$ which contains an abelian subgroup of finite index and has the property that the minimal polynomial of every $s \in S$ splits over $k$. Then $A \simeq M_n(k)$. 
Proof. Let $S_1$ denote the abelian subgroup of finite index in $S$. Observe that $S_1$ has a subgroup that is both normal and of finite index in $S$ (the kernel of the action of $S$ on the set of left cosets $S/S_1$) so there is no harm in assuming that $S_1$ is normal in $S$. Let $B$ be the $k$-algebra generated by $S_1$. By Clifford’s theorem (see [3]) $B$ is semisimple and since it is commutative it is isomorphic to a direct sum of $m$ copies of $k$ for some $m$. Let $\{e_1, \ldots, e_m\}$ be the set of minimal orthogonal idempotents in $B$ upon which $S$ acts transitively by conjugation and set $C = C_S(e_1)$ the centralizer of $e_1$ in $S$. It is clear that given $a \in S$ we have $e_1 a e_1 \neq 0$ precisely when $a \in C$. So the central simple algebra $e_1 A e_1$ is spanned by a center-by-finite group $Ce_1$ (a homomorphic image of $C$) which obviously has the property that the minimal polynomial of every element in $Ce_1$ splits over $k$ and the claim follows by the previous proposition.

From now on we suppose, unless stated otherwise, that $k = \mathbb{Q}_p$, i.e., $k$ is a non-archimedian locally compact field that contains $\mathbb{Q}$ as a dense subfield. We fix an absolute value $|\cdot|$ on $k$ and let $\mathcal{O}_k$ denote the ring of $p$-adic integers.

For most of the facts concerning algebraic groups we refer the reader to [2].

Theorem 3. Let $A$ be a central simple algebra over $k$. Suppose $A$ is spanned over $k$ by a compact group $S \subset A^*$ with the property that the minimal polynomial of every element in $S$ splits over $k$. Then $A \cong M_n(k)$ and $S$ contains a commutative normal subgroup of finite index.

Proof. Let $G$ be the Zariski closure of $S$ in $A^*$. Then $G$ is an algebraic group defined over $k$. The absolute irreducibility of $S$ clearly implies that the connected component of the unit $G^0$, if not trivial, is a reductive group defined over $k$. If $G^0$ is trivial, then $S$ is a finite group and the theorem follows, therefore we assume that $G^0$ is not trivial. Being reductive, it is an almost direct product of its central torus $T$ and a semisimple group $H = (G^0, G^0)$ where both $T$ and $H$ are defined over $k$.

We want to show that $H$ is trivial so we assume the contrary. Now the group $S_1 = S \cap G^0$ is also compact, therefore it is a Lie group over $k$ by Cartan’s theorem (see [12]). Its Zariski closure is precisely $G^0$. We know, that the Zariski closure of $(S_1, S_1)$ is $H$ and consequently it is the Zariski closure of $S_2 = S \cap H$. This latter group is also compact and we let $L_a(S_2)$ denote its Lie algebra in the sense of [12]. By [8, Prop. 3.4] $L_a(S_2)$ is a Lie ideal in $L_a(H(k)) = L(H)(k)$. But this is a semisimple Lie algebra which, together with Zariski density and compactness of $S_2$, implies $L_a(S_2) = L(H)(k)$. To see this assume $L(H)(k) = L_a(S_2) \oplus V$ where $V \neq 0$ is the unique ideal in $L(H)(k)$, complementary to $L_a(S_2)$. Consider the restriction to $V$ of the adjoint representation of $H(k)$. The conditions imply that $S_2$ is mapped to a discrete compact group, therefore finite, but on the other hand Zariski dense in a nontrivial semisimple group, a contradiction.
Since \( L_a(S_2) = L(H)(k) \) the group \( S_2 \) contains a subgroup that is open in \( H(k) \). But every semisimple group \( H \) over a \( p \)-adic field \( k \) contains a maximal \( k \)-torus that is anisotropic over \( k \) (see [8, Thm. 6.21]) and it follows that there are elements arbitrarily close to identity in \( H(k) \) that cannot be diagonalized over \( k \) and again we have a contradiction. Therefore \( H \) is trivial and subsequently \( S_1 \) is an abelian normal subgroup of finite index in \( S \) and the claim follows by Corollary 2.

**Remark 4.** If the field \( k \) in the previous theorem were a general \( p \)-adic field one could still infer that \( S_2 \) is a Lie group over \( \mathbb{Q}_p, \mathbb{Q}_p < k \). But in this case its Lie algebra \( L_a(S_2) \), which need not be trivial (see the example at the end of the paper), is just a \( \mathbb{Q}_p \)-subalgebra in \( L(H)(k) \). In fact the same argument as above shows that it is never a nontrivial ideal in \( L(H)(k) \).

Keeping the notation of the previous theorem we also have the following corollary.

**Corollary 5.** The exponent of the finite group \( S/S_1 \) divides the number of roots of unity in \( k \) and the same is true for \( G/G^0 \).

*Proof.* The statement is obvious if \( S_1 \) is trivial so we assume the contrary. Let \( m \) denote the number of roots of unity in \( k \). Now the first assertion follows by observing that given \( a \in S \) the power \( a^m \) belongs to domain of convergence of the log series. Therefore we have \( a^m = \exp(b), b \in L_a(S) = L_a(S_1) \) and \( a^m \in S_1 \) by the definition of \( S_1 \). The second assertion follows immediately from Zariski density of \( S \) in \( G \).

**Theorem 6.** Let \( A \) be a central simple algebra over \( k \). Suppose \( A \) is spanned over \( k \) by a multiplicative semigroup \( S \subset A \) with the property that the minimal polynomial of every element in \( S \) splits over \( k \). Then \( A \) is isomorphic to \( M_n(k) \).

*Proof.* Observe that we can view \( A \) as matrices over some finite complete extension \( K \) of \( k \). The semigroup \( S \) is then an absolutely irreducible matrix semigroup in \( M_n(K) \) with the property that the spectrum of every \( a \in S \) is contained in \( k \). This property is preserved under multiplication by scalars in \( k \), but also under norm closure, which follows from the continuity of spectrum, a consequence of Krasner’s Lemma (see [6]). So we may assume that \( S \) is a closed subset of \( A \) (and \( M_n(K) \)) and \( S = kS = \{ \lambda s; \lambda \in k, a \in S \} \). Now let \( r \) be the minimum of the ranks of the nonzero elements in \( S \). We claim that there is a rank \( r \) element \( a \in S \) that is not a nilpotent. Assume the contrary. Let \( b \in S \) have rank \( r \). Then the semigroup \( SbS \) is a semigroup of nilpotents and by Levitzky’s theorem (see also [1]) it generates a nontrivial nilpotent ideal in \( A \) but \( A \) is simple, hence the contradiction. Next we claim that there is an idempotent \( e \in S \) with rank \( r \). Let \( a \in S \) be a nonnilpotent element of this minimal rank. By dividing it with an appropriate element
of $k^*$ we may assume that all the eigenvalues lie in the unit ball and that some (say $m$) have absolute value 1. By replacing $a$ with some power of it we may further assume that all the eigenvalues with absolute value 1 belong to $1 + p o_k$, where $p$ is the characteristic of the residue class field. Let $a = a_s + a_n$ be the Jordan decomposition of $a$ with $a_n^k = 0$ for some $k$. Consider the sequence

$$a^{p^n} = \sum_{j=0}^{k-1} \binom{p^n}{j} a_s^{p^n-j} a_n^j$$

for $n$ large enough. It is easy to see that this sequence converges to an idempotent $e \in S$ of rank $m$, so $m = r$ by minimality of rank. Consider the semigroup $e Se$. It clearly satisfies the conditions of $S$ and it generates the central simple $k$-algebra $eAe$. The theorem will be proved provided we show $eAe \simeq M_r(k)$.

Replacing $eAe$ for $A$ we see that one can assume that $S \subset A$ is a closed semigroup in which every nonzero element is invertible and all the eigenvalues of an element in $S$ have equal absolute value. Consider the set $S_1 = \{ a \in S; |nr(a)| = 1 \}$ which is clearly also a closed semigroup in $A$. Observe that $a \in S$ belongs to $S_1$ precisely when all its eigenvalues lie in $o_k^*$ and so every element of $S_1$ is integral over $o_k$. Since $S_1$ generates $A$ as a vector space over $k$ we have to show that $S_1$ is a compact group and then apply the previous theorem.

Consider the $o_k$ algebra $\Gamma \subseteq A$, generated by $S_1$. Then by [1] every element of $\Gamma$ is integral over $o_k$. By [10, Thm. 10.3] $\Gamma$ is an $o_k$-order in $A$ and therefore contained in a maximal $o_k$-order $\Delta$. The latter is compact and so is $S_1 \subset \Delta$. The only idempotent in the compact semigroup $S_1$ is the identity, so by [11, Thm. 3.5] $S_1$ is a compact group and the proof is complete.

As a corollary we obtain the following result which is already implicit in [9].

**Corollary 7.** Let $A$ be a central simple algebra over $\mathbb{Q}$. Assume $A$ is spanned by a multiplicative semigroup $S \subset A$ with the property that the minimal polynomial of every element in $S$ splits over $\mathbb{Q}$. Then $A$ is isomorphic to $M_n(\mathbb{Q})$.

**Proof.** By the Hasse-Brauer-Albert-Noether theorem [10, Thm. 32.11] it suffices to show that $A_P = A \otimes \mathbb{Q}_P$ is trivial in the Brauer group $\text{Br}(\mathbb{Q}_P)$ for every prime $P$. If $P$ is finite, then $\mathbb{Q}_P = \mathbb{Q}_p$ for some rational prime $p$ and we apply the previous theorem for the image of $S$ under natural embedding. Since there is only one infinite prime $P$ over $\mathbb{Q}$, namely $\mathbb{Q}_P = \mathbb{R}$, we also have that $A_P$ is trivial in this case by the product formula.
3. Some examples and remarks.

Example 8. As our first example we take the group of matrices \( S \subset M_2(k) \), generated by \( S_1 = \{ \text{diag} (\alpha^2, \beta^2); \alpha, \beta \in \mathbb{O}_k^* \} \) and \( t \), where \( t \) is the matrix of the transposition of the standard basis of \( k^2 \). An easy computation shows that this is indeed a compact irreducible group with the desired property. This example can easily be generalized to matrices of order \( 2^n \).

The next example shows that, although there are good reasons to believe that both the theorems of this article hold for any \( p \)-adic field \( k \), the structure of the compact group \( S \) with the desired properties can be much more complicated in this general case.

Example 9. Let \( k \) be a \( p \)-adic field and \( n \) a fixed number. Then there exists a finite extension field \( l > k \) with the property that every polynomial of degree less or equal to \( n \) in \( k[X] \) splits over \( l \) (see \([8, \text{Prop. 6.13}]\)). Let \( A \) be a central simple algebra of reduced degree \( n \) over \( k \) and \( \Delta \subset A \) a maximal order in \( A \). Let \( S \) be the image of \( \Delta^* \) under natural embedding \( A \rightarrow A \otimes_k l \cong M_n(l) \), \( a \mapsto a \otimes 1 \). Clearly \( S \) is compact, it spans \( M_n(l) \) and the minimal polynomial of every \( s \in S \) splits over \( l \).

We conclude with a remark and some open questions.

Remark 10. Let \( k \) be either \( \mathbb{Q}_p \) or \( \mathbb{R} \). It is an interesting question whether there exists an absolutely irreducible group \( S \subset GL_n(k) \) such that the spectra of its elements lie in \( k \) and such that the Zariski closure of \( S \) is semisimple. A similar argument as in the proof of Theorem 3 shows that such a group is necessarily discrete and, consequently, its Zariski closure not \( k \)-anisotropic. It is known (see \([13]\)) that \( SL_2(k) \) contains a free Zariski dense subgroup \( S \) and one can check easily that the spectra of all its elements under any \( k \)-representation of \( SL_2 \) lie in \( k \). It is an open question what are the necessary and sufficient conditions on a semisimple group (for instance being \( k \)-split) in order for it to be the Zariski closure of a group \( S \) with the desired property.

References


Received September 9, 2001 and revised November 16, 2001.

UNIVERSITY OF LJUBLJANA
FACULTY OF MATHEMATICS AND PHYSICS
JADRANSKA 19
SI-1000 LJUBLJANA
SLOVENIA
E-mail address: janez.bernik@fmf.uni-lj.si
AMPLE FAMILIES, MULTIHOMOGENEOUS SPECTRA, 
AND ALGEBRAIZATION OF FORMAL SCHEMES

Holger Brenner and Stefan Schröer

Generalizing homogeneous spectra for rings graded by natural numbers, we introduce multihomogeneous spectra for rings graded by abelian groups. Such homogeneous spectra have the same completeness properties as their classical counterparts, but are possibly nonseparated. We relate them to ample families of invertible sheaves and simplicial toric varieties. As an application, we generalize Grothendieck’s Algebraization Theorem and show that formal schemes with certain ample families are algebraizable.

Introduction.

A powerful method to study algebraic varieties is to embed them, if possible, into some projective space $\mathbb{P}^n$. Such an embedding $X \subset \mathbb{P}^n$ allows you to view points $x \in X$ as homogeneous prime ideals in some $\mathbb{N}$-graded ring. The purpose of this paper is to extend this to divisorial varieties, which are not necessarily quasiprojective.

The notion of divisorial varieties is due to Borelli [2]. The class of divisorial varieties contains all quasiprojective schemes, smooth varieties, and locally $\mathbb{Q}$-factorial varieties. Roughly speaking, divisoriality means that there is a finite collection of invertible sheaves $\mathcal{L}_1, \ldots, \mathcal{L}_r$ so that the whole collection behaves like an ample invertible sheaf. Such collections are called ample families.

Our main idea is to define homogeneous spectra $\text{Proj}(S)$ for multigraded rings, that is, for rings graded by an abelian group of finite type. We obtain $\text{Proj}(S)$ by patching affine pieces $D_+(f) = \text{Spec}(S(f))$, where $f \in S$ are certain homogeneous elements. Roughly speaking, we demand that $f$ has many homogeneous divisors. Multihomogeneous spectra share many properties of classical homogeneous spectra. For example, they are universally closed, and the intersection of two affine open subsets is affine. They are, however, not necessarily separated.

Roberts [23] gave a similar construction for $\mathbb{N}^r$-graded rings $S$ satisfying certain conditions on homogeneous generators. He used it to study Hilbert functions in several variables and local multiplicities. Roberts’ homogeneous spectrum is an open subset of ours.
It turns out that a scheme is divisorial if and only if it admits an embedding into suitable multihomogeneous spectra. More precisely, we shall characterize ample families of invertible sheaves in terms of \( \text{Proj}(S) \) for the multigraded ring of global sections \( S = \bigoplus_{d \in \mathbb{N}^r} \Gamma(X, L_1^{d_1} \otimes \cdots \otimes L_r^{d_r}) \). Generalizing Grauert’s Criterion for ample sheaves, we characterize ample families also in terms of affine hulls and contractions. Furthermore, we give a cohomological characterization which is analogous to Serre’s Criterion for ample sheaves. We also relate homogeneous spectra to Cox’s homogeneous coordinate rings for toric varieties \( [6] \).

As an application, we shall generalize Grothendieck’s Algebraization Theorem. The result is that a proper formal scheme \( X \to \text{Spf}(R) \) is algebraizable if there is a finite collection of invertible formal sheaves restricting to an ample family on the closed fiber and satisfying an additional condition.

### 1. Grauert’s criterion for ample families.

In this section, we shall generalize Grauert’s criterion for ample sheaves to ample families. Given a collection of invertible sheaves \( L_1, \ldots, L_r \), we use multiindices and set \( L^d = L_1^{d_1} \otimes \cdots \otimes L_r^{d_r} \) for each \( d = (d_1, \ldots, d_r) \in \mathbb{Z}^r \). Let us start with the defining property of ample families:

**Proposition 1.1.** Let \( X \) be a quasicompact and quasiseparated scheme. For a family \( L_1, \ldots, L_r \) of invertible sheaves, the following are equivalent:

(i) The open sets \( X_f \) with \( f \in \Gamma(X, L^d) \) and \( d \in \mathbb{N}^r \) form a base of the topology.

(ii) For each \( x \in X \), there is some \( d \in \mathbb{N}^r \) and \( f \in \Gamma(X, L^d) \) so that \( X_f \) is an affine neighborhood of \( x \).

(iii) For each point \( x \in X \), there is a \( \mathbb{Q} \)-basis \( d_1, \ldots, d_r \in \mathbb{N}^r \) and global sections \( f_i \in \Gamma(X, L^{d_i}) \) so that \( X_{f_i} \) are affine neighborhoods of \( x \).

**Proof.** To see \((ii) \Rightarrow (iii)\), choose a degree \( d \in \mathbb{N}^r \) and a section \( f \in \Gamma(X, L^d) \) so that \( U = X_f \) is an affine neighborhood of the point \( x \). Choose a \( \mathbb{Q} \)-basis \( e_1, \ldots, e_r \in \mathbb{N}^r \). Since \( U \) is affine, we find sections \( f_i' \in \Gamma(U, L^{e_i}) \) satisfying \( f_i'(x) \neq 0 \). According to \([10]\) Theorem 6.8.1, the sections \( f_i = f_i' \otimes f^n \) extend from \( U \) to \( X \) for \( n \gg 0 \). The sections \( f \) and \( f_i \) have degrees \( d \) and \( d_i = nd + e_i \), respectively. Skipping one of them we have a basis and the desired sections. The other implications are clear. \( \square \)

Following Borelli \([2]\), we call a finite collection \( L_1, \ldots, L_r \) of invertible \( \mathcal{O}_X \)-modules an **ample family** if the scheme \( X \) is quasicompact and quasiseparated, and the equivalent conditions in Proposition 1.1 hold. A scheme is called **divisorial** if it admits an ample family of invertible sheaves.

Recall that a scheme is **separated** if the diagonal embedding \( X \subset X \times X \) is closed, and **quasiseparated** if the diagonal is quasicompact. Note that, in contrast to the definition of ample sheaves \((11) \text{ Def. 4.5.3}\), we do not require
separateness for ample families. However, the possible nonseparatedness is rather mild:

**Proposition 1.2.** The diagonal embedding of a divisorial scheme is affine.

*Proof.* Let $X$ be a divisorial scheme, $\mathcal{L}$ be an invertible sheaf, $f \in \Gamma(X, \mathcal{L})$ a global section, and $U = \text{Spec}(A)$ an affine open subset. Then $U \cap X_f = \text{Spec}(A_f)$ is affine. Since $X$ is covered by affine open subsets of the form $X_f$, this ensures that the diagonal embedding $X \subset X \times X$ is affine. \hfill \Box

Obviously, schemes admitting ample invertible sheaves are divisorial. The following gives another large class of divisorial schemes:

**Proposition 1.3.** Normal noetherian locally $\mathbb{Q}$-factorial schemes with affine diagonal are divisorial schemes.

*Proof.* As in [14] II 2.2.6, the complement of an affine dense open subset is a Weil divisor. By assumption it is $\mathbb{Q}$-Cartier, so $X$ is divisorial. Using quasi-compactness, we find finitely many effective Cartier divisors $D_1, \ldots, D_r \subset X$ with $\bigcap D_i = \emptyset$. By Proposition 1.1, the invertible sheaves $\mathcal{O}_X(D_i)$ form an ample family. \hfill \Box

Here are two useful properties of divisorial schemes:

**Proposition 1.4.** For divisorial noetherian schemes, the following hold:

(i) Each coherent $\mathcal{O}_X$-module admits a resolution with locally free $\mathcal{O}_X$-modules of finite rank.

(ii) There is a noetherian ring $A$ together with a smooth surjective affine morphism $\text{Spec}(A) \to X$.

*Proof.* The first assertion is due to Borelli [3] Theorem 3.3. The second statement is called the Jouanolou–Thomason trick. For a proof, see [24] Proposition 4.4. \hfill \Box

Grauert’s Criterion states that a line bundle is ample if and only if its zero section contracts to a point (see [9] p. 341 and [11] Theorem 8.9.1). The task now is to generalize this to families of line bundles. To do so, we shall use vector bundles. Recall that the category of locally free $\mathcal{O}_X$-modules $\mathcal{E}$ is antiequivalent to the category of vector bundles $\pi : B \to Z$ via $B = \text{Spec} S(\mathcal{E})$ and $\mathcal{E} = \pi^*(\mathcal{O}_B)_1$. Under this correspondence, the sections $f \in \Gamma(X, S(\mathcal{E}))$ correspond to functions $f \in \Gamma(B, \mathcal{O}_B)$. For $f \in \Gamma(X, S^n(\mathcal{E}))$ define $X_f = \{x \in X : f(x) \neq 0\}$. Then $B_f \subset \pi^{-1}(X_f)$ and $X_f = \pi(B_f)$.

A locally free sheaf $\mathcal{E}$ is called ample if the invertible sheaf $\mathcal{O}_P(1)$ is ample on $P = \mathbb{P}(\mathcal{E})$ (see [15]). This easily implies that the open subsets $X_f \subset X$ with $f \in \Gamma(X, S^n(\mathcal{E}))$ generate the topology. In contrast to line bundles, the latter condition is not sufficient for ampleness. Let us characterize this condition:
Theorem 1.5. Suppose $X$ is quasicompact and quasiseparated. Let $\pi : B \to X$ be a vector bundle, $E = \pi_*(\mathcal{O}_B)_1$ the corresponding locally free sheaf, and $Z \subset B$ the zero section. Then the following are equivalent:

(i) The open subsets $X_f \subset X$ with $f \in \Gamma(X, S^n(E))$ generate the topology.
(ii) For every point $x \in X$ there is a function $f \in \Gamma(B, \mathcal{O}_B)$ so that $B_f$ is affine and $\pi^{-1}(x) \cap B_f \neq \emptyset$.
(iii) There is a scheme $B'$, and a morphism $q : B \to B'$, and an open subset $U' \subset B'$ so that the following holds: The image $q(Z) \subset B'$ admits a quasiaffine open neighborhood, $B \to B'$ induces an isomorphism $q^{-1}(U') \cong U'$ and the projection $q^{-1}(U') \to X$ is surjective.
(iv) There exists an open subset $Z \subset W \subset B$ such that for every $x \in X$ there is a function $f \in \Gamma(W, \mathcal{O}_B)$ so that $W_f$ is affine and $W_f \cap \pi^{-1}(x)$ is nonempty.

Proof. We shall prove the implications $(i) \Rightarrow \cdots \Rightarrow (iv) \Rightarrow (i)$. First assume $(i)$. Fix a point $x \in X$. Choose an affine neighborhood $x \in V$ and a section $f \in \Gamma(X, S^n(E))$ so that $x \in X_f \subset V$. Then $B_f = \pi^{-1}(V)_f$ is affine and $\pi^{-1}(x) \cap B_f \neq \emptyset$.

Assume $(ii)$ holds. Set $B' = B^\text{aff} = \text{Spec } \Gamma(B, \mathcal{O}_B)$. According to [10] Corollary 6.8.3, the map $\Gamma(B, \mathcal{O}_B)_f \to \Gamma(B_f, \mathcal{O}_B)$ is bijective, so the affine hull $q : B \to B^\text{aff}$ induces an isomorphism $B_f = q^{-1}(D(f)) \to D(f)$ for $B_f$ affine. Then take $U' = \bigcup_f D(f)$ and $U = \bigcup_f B_f$, where the union runs over $f$, $B_f$ affine.

Assume that $(iii)$ holds. Let $W' \subset B'$ be a quasiaffine neighborhood of $q(Z)$. Then $W = q^{-1}(W')$ is an open neighborhood of the zero section. Fix a point $x \in X$ and let $\eta \in \pi^{-1}(x)$ be the generic point, such that $\eta \in q^{-1}(U') \cap W$. Then we find $f \in \Gamma(W', \mathcal{O}_B)$ so that $q(\eta) \in W'_f \subset U' \cap W'$ is affine and $\eta \in W_f = q^{-1}(W'_f) \cong W'_f$ is an affine neighborhood.

Now suppose $(iv)$ holds. Let $x \in V \subset X$. Then there exists $f \in \Gamma(W, \mathcal{O}_B)$ so that $W_f \subset \pi^{-1}(V) \cap W$ is affine and $\pi^{-1}(x) \cap W_f \neq \emptyset$. Let $\mathcal{I} \subset \mathcal{O}_W$ be the ideal of the zero section $Z \subset W$. Then $\mathcal{O}_W/\mathcal{I}^{n+1} = \bigoplus_{d=0}^n \mathcal{O}^d(E)$, and $f \in \Gamma(W, \mathcal{O}_B)$ has a Taylor series expansion $f = \sum_{d=0}^\infty f_d$ with $f_d \in \Gamma(X, \mathcal{O}^d(E))$. Choose a degree $d \geq 0$ with $f_d(x) \neq 0$. Then $x \in X_{f_d} \subset V$ is the desired open neighborhood of $x$.

\begin{remark}
If $X$ is connected and proper over a base field, the image of the zero section $Z \subset B$ in any quasiaffine scheme is a closed point. In this case, the assumption in Condition $(iii)$ implies that $q$ contracts $Z \cong X$ to a point.
\end{remark}

The preceding result yields a characterization of ample families $\mathcal{L}_1, \ldots, \mathcal{L}_r$ in terms of the corresponding locally free sheaf $\mathcal{E} = \bigoplus \mathcal{L}_i$.

Corollary 1.7. Suppose $X$ is quasicompact and quasiseparated. A family of invertible sheaves $\mathcal{L}_1, \ldots, \mathcal{L}_r$ is ample if and only if the vector bundle
\[ \pi : B \to X \text{ with } \pi_*(\mathcal{O}_B)_1 = \mathcal{L}_1 \oplus \cdots \oplus \mathcal{L}_r \text{ satisfies the equivalent conditions in Theorem 1.5.} \]

**Proof.** If the family is ample, then Condition (i) of Theorem 1.5 holds. For the converse, fix a point \( x \in X \), and choose a section \( f \in \Gamma(X, S^n(\mathcal{E})) \) so that \( x \in X_f \subset V \) where \( V \) is affine. Write \( f = \sum f_d \) according to the decomposition \( S^n(\mathcal{E}) = \bigoplus d \mathcal{L}^d \), where the sum runs over all degrees \( d \in \mathbb{N}^r \) with \( n = \sum d_i \). Pick a summand \( f_d \) with \( f_d(x) \neq 0 \). Then \( X_{f_d} \subset X_f \subset V \) is the desired affine open neighborhood. \( \square \)

**Remark 1.8.** In the situation of 1.7, the vector bundle \( B = B_1 \times_X \cdots \times_X B_r \) decomposes into \( B_i = \text{Spec } S(\mathcal{L}_i) \). The affine hull \( B \to B^\text{aff} \) is an isomorphism outside the coordinate hyperplanes. The following examples illustrate what may happen on the union of the coordinate hyperplanes:

**Example 1.9.** Set \( X = \mathbb{P}^1 \times \mathbb{P}^1 \). First, let \( \mathcal{L}_1 \) and \( \mathcal{L}_2 \) be invertible sheaves of bidegree \((1,0)\) and \((0,1)\), respectively. This is an ample family because \( \mathcal{L}_1 \otimes \mathcal{L}_2 \) is ample. Set \( B_1 = \text{Spec } S(\mathcal{L}_1) \) and consider the corresponding rank two vector bundle \( B = B_1 \times_X B_2 \). On each summand \( B_i \), the affine hull \( B \to B^\text{aff} \) restricts to the morphism \( B_i = \mathbb{A}^2 \times \mathbb{P}^1 \overset{pr_1}{\to} \mathbb{A}^2 \overset{g}{\to} \mathbb{A}^2 \) where \( g : \mathbb{A}^2 \to \mathbb{A}^2 \) is the blowing-up of the origin.

Now let \( \mathcal{L}_1 \) and \( \mathcal{L}_2 \) be invertible sheaves of bidegree \((2,-1)\) and \((-1,1)\), respectively. This is an ample family, because \( \mathcal{L}_1^2 \otimes \mathcal{L}_2^3 \) is an ample sheaf. Here the affine hull \( B \to B^\text{aff} \) contracts the union of the coordinate hyperplane \( B_1 \cup B_2 \) to a point.

In the next examples we consider the following situation: Let \( k \) denote a field and let \( A \) be an \( \mathbb{N} \)-graded \( k \)-algebra of finite type with \( A = k[A_1] \). Let \( n_1, \ldots, n_r \in \mathbb{Z} \) and consider the family \( \mathcal{O}_X(n_1), \ldots, \mathcal{O}_X(n_r) \) on \( X \). Now we describe the corresponding vector bundle and its ring of global sections. (The second description uses Proj of a \( \mathbb{Z} \)-graded ring, which we introduce in the next section.)

**Proposition 1.10.** Let \( A, X \) and \( n_1, \ldots, n_r \) as above. Let \( S = A[T_1, \ldots, T_r] \) be \( \mathbb{Z} \)-graded by degrees \( \deg(T_j) = -n_j \). Then the vector bundle \( B = \text{Spec } S(\mathcal{O}_X(n_1) \oplus \cdots \oplus \mathcal{O}_X(n_r)) \) is

\[ B \cong \bigcup_{f \in A_1} D_+(f) \subset \text{Proj } S_{\geq 0} \quad \text{and} \quad B \cong \bigcup_{f \in A_1} D_+(f) \subset \text{Proj } S. \]

If \( A \) is normal and \( \dim X \geq 1 \), then \( \Gamma(B, \mathcal{O}_B) = S_0 \).

**Proof.** Let \( R = S \) or \( S_{\geq 0} \). \( R_0 \) has also a \( \mathbb{Z}^r \)-gradation with \( (R_0)_d = \{aT^d : a \in A_{d_1, n_1 + \cdots + d_r n_r}\} \), \( d = (d_1, \ldots, d_r) \). There is a natural rational mapping \( \text{Proj } R \to \text{Proj } A \) which is defined on \( \bigcup D_+(f) \), where the union
runs over \( f \in A_1 \). This is an affine morphism, thus we can check the identities by looking at the rings of global sections and at the restriction maps. We have

\[
\Gamma(\pi^{-1}(D_+(f)), \mathcal{O}_B)_d = \Gamma(D_+(f), \mathcal{O}_X(n_1) \oplus \cdots \oplus \mathcal{O}_X(n_r))_d \\
= \Gamma(D_+(f), \mathcal{O}_X(d_1 n_1 + \cdots + d_r n_r)) \\
= (A_f)_{d_1 n_1 + \cdots + d_r n_r} \\
= ((R_f)_0)_d \\
= \Gamma(D_+(f), \mathcal{O}_\text{Proj} R)_d
\]

and the restriction maps respect these identities. The last statement follows if we replace \( D_+(f) \) by \( U = \bigcup_{f \in A_1} D_+(f) \). Then

\[
\Gamma(U, \mathcal{O}_X(d_1 n_1 + \cdots + d_r n_r)) = \Gamma(D(A_1), (\mathcal{O}_\text{Spec}(A))_{d_1 n_1 + \cdots + d_r n_r}) \\
= A_{d_1 n_1 + \cdots + d_r n_r},
\]

since \( A \) is normal and \( V(A_1) \) has codimension \( \geq 2 \). \qed

**Example 1.11.** Let \( X = \mathbb{P}^m \) be the homogeneous spectrum of \( k[X_0, \ldots, X_m] \), \( m \geq 1 \). Let \( \mathcal{L}_1 = \mathcal{O}_X(1) \) and \( \mathcal{L}_2 = \mathcal{O}_X(-1) \). Then \( B_1 = \text{Spec} S(\mathcal{O}(1)) \) is the blowing up of the vertex point in \( k^{m+1} \) and \( B_2 = \text{Spec} S(\mathcal{O}(-1)) \) is the projection from a point in \( \mathbb{P}^{m+1} \). The ring of global sections of the rank two vector bundle \( B = B_1 \times_X B_2 \) is the polynomial algebra

\[
\Gamma(B, \mathcal{O}_B) = k[X_0 S, \ldots, X_m S, ST] \subset k[X_0, \ldots, X_m][S, T]
\]

where \( S, T \) are indeterminates with degrees \( \deg(S) = -1 \) and \( \deg(T) = 1 \). The affine hull \( B \rightarrow B^{\text{aff}} \) contracts \( B_2 \) to a point and is an isomorphism on the complement \( B - B_2 \). We have

\[
\Gamma(B, D_+(X_i), \mathcal{O}_B) = k[X_0/X_i, \ldots, X_m/X_i, X_iS, T/X_i],
\]

and the affine hull is given by

\[
X_j S \mapsto \frac{X_j}{X_i} X_i S \quad \text{and} \quad ST \mapsto \frac{T}{X_i} X_i S.
\]

**Example 1.12.** Again let \( A = k[X_0, \ldots, X_m] \) (\( m \geq 1 \)) and consider the family \( \mathcal{L}_1 = \mathcal{O}_X(1), \ldots, \mathcal{L}_r = \mathcal{O}_X(1) \). Then \( \Gamma(B, \mathcal{O}_B) \) is the determinantal algebra

\[
k[X_i T_j \mid 0 \leq i \leq m, 1 \leq j \leq r] \subset k[X_0, \ldots, X_m, T_1, \ldots, T_r],
\]

generated by the entries of the \((m+1) \times r\)-matrix \((X_i T_j)\), with relations given by the \(2 \times 2\)-minors \((X_i T_j)(X_k T_l) - (X_i T_l)(X_k T_j)\). The affine hull \( \varphi : B \rightarrow B^{\text{aff}} \) contracts exactly the zero section to a point. The affine hull \( \varphi \) may
also be described as the blowing-up of the column ideal \((X_0T_1, \ldots, X_mT_1)\). For this blowing up is given by \(\text{Proj } k[X_iT_j][X_0T_1U, \ldots, X_mT_1U]\), and

\[
k[X_iT_j][X_0T_1U, \ldots, X_mT_1U] \cong k[X_iT_j][X_0, \ldots, X_m]
\]

\[
= k[X_0, \ldots, X_m, T_1, \ldots, T_r]_{\geq 0}.
\]

2. The homogeneous spectrum of a multigraded ring.

Generalizing the classical notion of homogeneous coordinates, Grothendieck defined homogeneous spectra for \(\mathbb{N}\)-graded rings ([11] §2). In this section, we shall generalize his approach to multigraded rings. Let \(D\) be a finitely generated abelian group and let \(S = \bigoplus_{d \in D} S_d\) be a \(D\)-graded ring. Note that by [13] I 4.7.3, such gradings correspond to actions of the diagonalizable group scheme \(\text{Spec}(S_0[D])\) on the affine scheme \(\text{Spec}(S)\).

According to geometric invariant theory (see [22] Thm. 1.1), the projection \(\text{Spec}(S) \to \text{Spec}(S_0)\) is a categorical quotient in the category of schemes. There is a quotient \(\text{Spec}(S) \to \text{Quot}(S)\) in the category of ringed spaces as well. In general, the latter is quite different from the first. However, we have the following favorable situation: Call the ring \(S\) periodic if the degrees of the homogeneous units \(f \in S^\times\) form a subgroup \(D' \subset D\) of finite index. In this case, we may choose a free subgroup \(D' \subset D\) of finite index, such that \(S' = \bigoplus_{d \in D'} S_d\) is a Laurent polynomial algebra \(S_0[T_1^{\pm 1}, \ldots, T_r^{\pm 1}]\).

**Lemma 2.1.** For periodic rings \(S\), the projection \(\text{Spec}(S) \to \text{Spec}(S_0)\) is a geometric quotient in the sense of geometric invariant theory.

**Proof.** Choose a free subgroup \(D' \subset D\) of finite index as above. The corresponding inclusion of the Veronese subring \(S' \subset S\) is an integral ring extension, because \(D/D'\) is torsion, such that \(\text{Spec}(S) \to \text{Spec}(S')\) is a closed morphism. By [22] Amplification 1.3, this morphism is a geometric quotient.

Since \(S'\) is a Laurent polynomial ring, \(\text{Spec}(S')\) is a principal homogeneous space for the induced action of \(\text{Spec}(S_0[\mathbb{Z}^r])\) and the projection \(\text{Spec}(S') \to \text{Spec}(S_0)\) is a geometric quotient. Being the composition of two geometric quotients, \(\text{Spec}(S) \to \text{Spec}(S_0)\) is a geometric quotient as well. \(\square\)

In light of this, we seek to pass from a given graded ring \(S\) to periodic rings via localization. An element \(f \in S\) is called relevant if it is homogeneous and the localization \(S_f\) is periodic. Equivalently, the degrees of all homogeneous divisors \(gf^n\), \(n \geq 0\) generate a subgroup \(D' \subset D\) of finite index. Since geometric quotients are quotients in the category of ringed spaces, localization of relevant elements yields open subschemes

\[
D_+(f) = \text{Spec}(S_f) \subset \text{Quot}(S)
\]
inside the ringed space Quot(S). Here \( S_f \subset S_f \) is the degree zero part of the localization. This leads to the following definition:

**Definition 2.2.** Let \( D \) be a finitely generated abelian group and let \( S = \bigoplus_{d \in D} S_d \) be a \( D \)-graded ring. We define the scheme

\[
\text{Proj}(S) = \bigcup_{f \in S \text{ relevant}} D_+(f) \subset \text{Quot}(S),
\]

and call it the **homogeneous spectrum** of the graded ring \( S \).

For \( \mathbb{N} \)-gradings, this coincides with the usual definition. As in the classical situation, we define \( S_+ \subset S \) to be the ideal generated by all relevant \( f \in S \). The corresponding invariant closed subscheme \( V(S_+) \) is called the **irrelevant subscheme**. The complementary invariant open subset \( \text{Spec}(S) - V(S_+) \) is called the **relevant locus**. Obviously, we obtain an affine projection

\[
\text{Spec}(S) - V(S_+) \longrightarrow \text{Proj}(S),
\]

which is a geometric quotient for the induced action.

**Remark 2.3.** The points \( x \in \text{Proj}(S) \) correspond to graded (not necessarily prime) ideals \( p \subset S \) not containing \( S_+ \) such that the subset of homogeneous elements \( H \subset S - p \) is closed under multiplication. The stalk of the structure sheaf at \( x \in \text{Proj}(S) \) is canonically isomorphic to \( (H^{-1}S)_0 \).

To proceed, we need a finiteness condition for multigraded rings. In the special case \( D = \mathbb{Z} \), the following is due to Bruns and Herzog ([5] Thm. 1.5.5):

**Lemma 2.4.** Let \( S \) be a ring graded by a finitely generated abelian group \( D \). Then the following are equivalent:

(i) The homogeneous ideals of \( S \) satisfy the ascending chain condition.

(ii) The ring \( S \) is noetherian.

(iii) \( S_0 \) is noetherian and \( S \) is an \( S_0 \)-algebra of finite type.

If \( S \) is noetherian and \( M \subset D \) is a finitely generated submonoid, then \( S_M \) is also noetherian.

**Proof.** The implications (iii) \( \Rightarrow \) (ii) and (ii) \( \Rightarrow \) (i) are trivial, so assume that (i) holds. We start with some preparations. For a submonoid \( M \subset D \), let \( S_M = \bigoplus_{d \in M} S_d \) be the corresponding Veronese subring. Let \( D' \subset D \) denote a subgroup. Then the ring \( S_{D'} \subset S \) is a direct summand. Therefore \( S_{D'} \) satisfies also the ascending chain condition for homogeneous ideals. In particular, \( S_0 \) is noetherian. It follows at once that \( S_e \) are noetherian \( S_0 \)-modules. Let \( D \cong D' \oplus T \), where \( T \) is finite and \( D' \) free. Then \( S \) is finite over \( S_{D'} \) and \( S_D \) fulfills (i). Thus we may assume that \( D \) is free. Let us call a free submonoid \( M \subset D \) a **quadrant** if \( M = \bigoplus \mathbb{N}d_i \), where the \( d_i \in D \) is a subset of a \( \mathbb{Z} \)-basis for \( D \).
Claim. For each quadrant $M \subset D$ and each degree $e \in D$, the $S_M$-module $S_{M+e} = \bigoplus_{d \in M} S_{d+e}$ is finitely generated.

We prove this by induction on $r = \text{rank } M$. Since $S_e$ are noetherian $S_0$-modules, this holds for $r = 0$. Fix a quadrant $M$ of rank $r$ and suppose the claim is true for each quadrant of rank $r - 1$. Choose $d_1, \ldots, d_r \in D$ with $M = \bigoplus_{i=1}^r Nd_i$, and let $M_i = \bigoplus_{j \neq i} Nd_j$ be the $i$-th boundary quadrant.

Condition (i) implies that the graded $S$-ideal $S\cdot S_{M+e}$ is finitely generated. Choose homogeneous generators $f_1, \ldots, f_s \in S_{M+e}$. Then $S_d \subset \sum S_M f_i$ for each $d \in M + e$ with $d \geq \max \{\text{deg}(f_1), \ldots, \text{deg}(f_s)\}$. Fix such a degree $d \in M + e$. Clearly, there are finitely many $d_{ij} \in M + e$ with

$$M + e = (M + d) + \sum_i \sum_j (M_i + d_{ij}).$$

By induction, each $S_{M_i+d_{ij}}$ is a finitely generated $S_{M_i}$-module, and we conclude that $S_{M+e}$ is a finitely generated $S_M$-module. This proves the claim.

Fix a quadrant $M = \sum Nd_i$ and set $M^* = M - 0 = \sum (M + d_i)$. Then $S_{M+d_i}$ is a finitely generated $S_M$-module and thus $S_{M^*} = \sum S_{M+d_i}$ is finitely generated.

Let $S_{M^*} = (f_1, \ldots, f_n)$, where $f_i$ are homogeneous of degree $> 0$. We show by induction on $M \cong \mathbb{N}^r$ that $S_M = S_0[f_1, \ldots, f_n]$. Let $g \in S_{d}, d \in M$ and suppose that $S_e \subset S_0[f_1, \ldots, f_n]$ for all $e < d$. We have $g = h_1 f_1 + \cdots + h_n f_n$, where $\text{deg}(h_i) = \text{deg}(g) - \text{deg}(f_j) < d$ and the result follows.

Since $D$ is a finite union of quadrants with $\text{rank}(M) = \text{rank}(D)$, we conclude that $S$ is a finitely generated $S_0$-algebra.

To prove the additional statement, let $g_i, i \in I$ be a generating system for $S$ with degrees $d_i$. We may assume that $M$ is saturated, thus we may describe $M$ with finitely many linear forms $\psi_k : D \rightarrow \mathbb{Z}$, say $M = \bigcap_k \psi_k^{-1}(\mathbb{N})$. Consider the mapping $\varphi : \mathbb{N}^I \rightarrow D, (n_i)_{i \in I} \mapsto \sum_i n_i d_i$. Then $\varphi^{-1}(M) = \bigcap_k (\psi_k \circ \varphi)^{-1}(\mathbb{N})$ and therefore $\varphi^{-1}(M)$ is finitely generated, say by $r_j, j \in J$. Set $e_j = \varphi(r_j)$ in $M$. We claim that $S_M$ is generated by elements of degree $e_j, j \in J$. An element $f \in S_d, d \in M$ can be written as the sum of products $\prod_{i \in I} g_i^{n_i}$, where $d = \sum_{i \in I} n_i d_i$. But then $(n_i)_{i \in I} = \sum_j m_j r_j$ and

$$\prod_{i \in I} g_i^{n_i} = \prod_{i \in I} g_i^{\sum_j m_j r_j} = \prod_{j \in J} \left( \prod_{i \in I} g_i^{r_j} \right)^{m_j}.$$

Thus $\prod g_i^{n_i}$ is a product of elements in $S_{e_j}$. \hfill \Box

We have the following finiteness condition for homogeneous spectra:

**Proposition 2.5.** The morphism $\text{Proj}(S) \rightarrow \text{Spec}(S_0)$ is universally closed and of finite type, provided that $S$ is noetherian.
Proof. By Lemma 2.4, the ring $S_0$ is noetherian and the $S_0$-algebra $S$ is of finite type. The relevant locus $\text{Spec}(S) - V(S_+)$ is quasicompact and surjects onto $\text{Proj}(S)$, so the homogeneous spectrum is quasicompact.

Next, we check that the projection $\text{Proj}(S) \to \text{Spec}(S_0)$ is locally of finite type. To do so, fix a relevant element $f \in S$, and let $D' \subset D$ be a free subgroup of finite index such that for each $d \in D'$, there is a homogeneous unit $g \in S_f^\times$ with $\text{deg}(g) = d$. Let $S' = \bigoplus_{d \in D'} S_d$ be the corresponding Veronese subring. By Lemma 2.4, the ring extension $S' \subset S$ is finite, so Artin–Tate [1] tells us that the ring extension $S_0 \subset S'$ is of finite type. Clearly, the localization $S'_f$ is isomorphic to a Laurent polynomial ring $S(f)[T_{i \pm 1}^\pm, \ldots, T_r^\pm]$. Setting $T_i = 1$, we deduce that $S_0 \subset S(f)$ is of finite type.

Finally, we verify universal closedness. Let $S_0 \to R_0$ be a base change and set $R = S \otimes_{S_0} R_0$. A direct argument gives $R_+ = S_+ \otimes_{S_0} R_0$. Hence we have to check that $h : \text{Proj}(S) \to \text{Spec}(S_0)$ is closed under the new hypothesis that $S$ is an $S_0$-algebra of finite type, that $S_+ \subset S$ is an ideal of finite type, and that each $S_d$ is a finitely generated $S_0$-module. Each closed subset of $\text{Proj}(S)$ is of the form $\text{Proj}(S/b) = V_+(b)$ for some graded ideal $b \subset S$, and we have $h(V_+(b)) \subset V(b_0)$. Consequently, it suffices to show that $h : \text{Proj}(S) \to \text{Spec}(S_0)$ has closed image.

Fix a point $x \in \text{Spec}(S_0)$ with $h^{-1}(x) = \emptyset$, and let $\mathfrak{p} \subset S_0$ be the corresponding prime ideal. We have to construct $g \in S_0 - \mathfrak{p}$ with $h^{-1}(\text{Spec}(S_0)_g) = \emptyset$. The condition $h^{-1}(x) = \emptyset$ signifies that each relevant $f \in S$ is nilpotent in $S/\mathfrak{p}S$. Hence $S_{k}^\times \subset \mathfrak{p}S_+$ for some integer $k > 0$. Choose finitely many homogeneous $g_1, \ldots, g_n \in S$ with $S = S_0[g_1, \ldots, g_n]$, and set $d_i = \text{deg}(g_i)$. Call a degree $d \in D$ generic if for each linear combination $d = \sum n_i d_i$ with nonnegative coefficients, the set $\{d_i \mid n_i \geq k\}$ generates a subgroup of finite index. Then the set of nongeneric degrees is a union of a finite set and finitely many affine hyperplanes.

Let $d = \sum n_i d_i$ be generic and consider $g = \prod g_i^{n_i} \in S_d$. Then one may write $g = g_1^{k_1} \cdots g_r^{k_r} \cdot g'$ such that $g_1 \cdots g_r$ is relevant. Using $S_{k_i}^\times \subset \mathfrak{p}S_+$, we see $g \in \mathfrak{p}S_d$. This gives $\mathfrak{p}S_d = S_d$ for all generic degrees $d \in D$, and the Nakayama Lemma gives $(S_d)_\mathfrak{p} = 0$.

Next, choose finitely many relevant $f_1, \ldots, f_m \in S$ with $S_+ = \sqrt{(f_1, \ldots, f_m)}$. We may assume $f_i = f_{i_1} \cdots f_{i_r}$ such that each sequence $\deg(f_{i_1}), \ldots, \deg(f_{i_r}) \in D$ generates a subgroup of finite index. For each $f_i$, choose a linear combination with positive coefficients $e_i = \sum_j n_j \deg(f_{ij}) \in D$ that is generic. Then some $g \in S_0 - \mathfrak{p}$ annihilates all $S_{e_i}$. Consequently, each $f_i \in S[1/g]$ is nilpotent, hence the preimage $h^{-1}(\text{Spec}(S_0)_g)$ is empty. □

Remark 2.6. Roberts ([23], Sect. 8.2) introduced multihomogeneous spectra for certain $\mathbb{N}^r$-graded rings. To explain Roberts' conditions, let $S_{[j]} \subset S$, $0 \leq j \leq r$ be the graded subring generated by all homogeneous elements whose degrees are of the form $(i_1, \ldots, i_j, 0, \ldots, 0)$. Then Roberts assumes
that each $S_{[j+1]}$ is generated over $S_{[j]}$ by finitely many elements of degree $(i_1, \ldots, i_j, 1, 0, \ldots, 0)$. Now Roberts’ homogeneous spectrum is the subset $\bigcup D_+(f) \subset \Proj(S)$, where the union runs over all $f \in S$ admitting a factorization $f = g_1 \cdots g_r$ so that $\deg(g_j)$ has the form $(i_1, \ldots, i_{j-1}, 1, 0, \ldots, 0)$, compare [23], Proposition 8.2.5.

3. Separation criteria.

In Proposition 2.5, we may say that $\Proj(S)$ is a complete $S_0$-scheme. We cannot, however, infer that it is proper. For example, set $S = k[X,Y]$ with degrees in $D = \mathbb{Z}$ given by $\deg(X) = 1$ and $\deg(Y) = -1$. Then $S_0 = k[XY]$ yields the affine line $\mathbb{A}^1_k$, and $\Proj(S)$ is the affine line with double origin, which is nonseparated. However, the following holds:

**Proposition 3.1.** The diagonal embedding of a homogeneous spectrum is affine.

*Proof.* The intersection $D_+(f) \cap D_+(g) = D_+(fg)$ is affine. □

Here is a criterion for separatedness, which trivially holds for $\mathbb{N}$-gradings:

**Proposition 3.2.** If for each pair $x, y \in \Proj(S)$ there is a relevant $f \in S$ with $x, y \in D_+(f)$, then $\Proj(S)$ is separated.

*Proof.* Under the assumption the affine open subsets $U \times U$, where $U \subset \Proj(S)$ is affine, cover $\Proj(S) \times \Proj(S)$. Clearly, the diagonal embedding of $\Proj(S)$ is closed over each of these open subsets, hence it is closed. □

The next task is to recognize large separated open subsets in $\Proj(S)$. Given a homogeneous element $f \in S$, let $H_f \subset S$ be the set of homogeneous divisors $g|f^n$, $n \geq 0$, and $C_f \subset D \otimes \mathbb{R}$ the closed convex cone generated by the degrees $\deg(g)$, $g \in H_f$. Note that a homogeneous element is relevant if and only if its cone has nonempty interior.

**Proposition 3.3.** Let $f_i \in S$ be a collection of relevant elements so that each closed convex cone $C_{f_i} \cap C_{f_j} \subset D \otimes \mathbb{R}$ has nonempty interior. Then $\bigcup D_+(f_i) \subset \Proj(S)$ is a separated open subset.

*Proof.* According to [10] Proposition 5.3.6, it suffices to show that the multiplication map $S(f_i) \otimes S(g) \rightarrow S(fg)$ is surjective for each pair of relevant elements $f, g \in S$ such that $C_f \cap C_g$ has nonempty interior. Note that, for each factorization $g^n = g_1 \cdots g_m$, we may replace $g$ by $g_1^{n_1} \cdots g_m^{n_m}$, $n_i > 0$ without changing the localization $S(f)$. Thus we may assume $\deg(g) \in C_f$. Passing to a suitable power of $g$, we may assume $\deg(g) = \sum n_i \deg(f_i)$, $n_i \geq 0$ with $f_i \in H_f$. Each element in $S(fg)$ has the form $a/(fg)^k$ with
$a \in S$ homogeneous, so
\[ \frac{a}{(fg)^k} = \frac{a}{f^k \prod f_i^{kn_i}} \cdot \frac{\prod f_i^{kn_i}}{g^k} \]
is contained in the image of $S(f) \otimes S(g)$. \hfill \Box

Next, we shall relate homogeneous spectra of multigraded polynomial algebras to toric varieties. Fix a ground ring $R$ and a free abelian group $M$ of finite rank. A simplicial torus embedding of the torus $T = \text{Spec}(R[M])$ is an equivariant open embedding $T \subset X$ that is locally given by $R[M \cap \sigma^\vee] \subset R[M]$ for some strongly convex, simplicial cone $\sigma \subset N_R$ in the dual lattice $N = \text{Hom}(M, \mathbb{Z})$. Here simplicial cone means that the cone is generated by a linear independent set. In contrast to the usual definition, we do not require that our torus embeddings are separated.

Simplicial torus embeddings occur in the following context, which is related to a construction of Cox [6]: Let $S = R[T_1, \ldots, T_k]$ be a $D$-graded polynomial algebra, such that the grading is given by a linear map $d : \mathbb{Z}^k \to D$ sending the $i$-th base vector to $\deg(T_i) \in D$. Let $M \subset \mathbb{Z}^k$ be the kernel.

**Proposition 3.4.** Notation as above. Then $\text{Proj}(S)$ is a (possibly nonseparated) simplicial torus embedding of the torus $\text{Spec}(R[M])$.

**Proof.** Let $I = \{1, \ldots, k\}$ be the index set for the indeterminates. Fix a relevant monomial $T^n = T_1^{n_1} \cdots T_k^{n_k}$ and let $J = \{i \in I \mid n_i > 0\}$ be its support. A direct argument gives $S(T^n) = R[M_J]$ for the monoid
\[ M_J = (\mathbb{Z}^J \oplus \mathbb{N}^{I-J}) \cap M \subset \mathbb{Z}^k. \]
Clearly, the submonoid $M_J \subset M$ is the intersection of $\text{Card}(I - J) \leq \text{rank}(M)$ half spaces. Therefore, its dual cone $\sigma \subset N_R$ is simplicial.

It remains to check $M = M_J + (-M_J)$. Let $D' \subset D$ be the subgroup generated by $\deg(T_i)$ with $i \in J$, and $m = \text{ord}(D/D')$ be its index. Then there are integers $\lambda_j \in \mathbb{Z}$, $j \in J$ solving the equation $\sum_{i \in J} \lambda_i \deg(T_i) = -m \sum_{i \in I-J} \deg(T_i)$. So the element $g \in \mathbb{Z}^k$ defined by
\[ g_i = \begin{cases} \lambda_i & \text{for } i \in J \\ m & \text{for } i \in I - J \end{cases} \]
lies in $M_J$. For each $f \in M$, we have $f + ng \in M_J$ for $n \gg 0$, hence $f = (f + ng) - ng$ is contained in $M_J + (-M_J)$. \hfill \Box

**Corollary 3.5.** If $S$ is finitely generated as $S_0$-algebra, then $\text{Proj}(S)$ is divisorial.

**Proof.** We may choose a $D$-graded polynomial $R$-algebra $S'$ with $S'_0 = S_0$, together with a graded surjection $S' \rightarrow S$. This induces a closed embedding
Proj($S$) ⊂ Proj($S'$), because for every relevant element $f \in S$ we may find a relevant element $f' \in S'$ mapping to it.

By Proposition 3.4, the scheme Proj($S'$) is a simplicial torus embedding, which has affine diagonal by Proposition 3.1, hence by Proposition 1.3 it is a divisorial scheme. Consequently, the closed subscheme Proj($S$) is divisorial as well.

**Corollary 3.6.** Suppose that $S$ is finitely generated over $S_0$. If each finite subset of Proj($S$) admits an affine neighborhood, then Proj($S$) is projective.

**Proof.** We already know that Proj($S$) is of finite type, universally closed, separated, and divisorial (Prop. 2.5, Prop. 3.2, and Cor. 3.5). Since each finite subset admits an affine neighborhood, the generalized Chevalley Conjecture ([18], Thm. 3) applies, and we conclude that Proj($S$) → Spec($S_0$) is projective. □

**Remark 3.7.** Let us make the torus embedding in Proposition 3.4 more explicit. For each subset $J \subset I = \{1, \ldots, k\}$, let $\sigma_J \subset \mathbb{R}$ be the convex cone generated by the projections $\text{pr}_i : \mathbb{Z}^k \to \mathbb{Z}$ restricted to $M$, $i \in J$. You easily check that $J \mapsto \sigma_{I - J}$ gives a bijection between the subsets $J \subset I$ with $\prod_{j \in J} T_j$ relevant, and the strongly convex simplicial cones $\sigma_{I - J} \subset \mathbb{R}$. Let us call such subsets relevant. Then the torus embedding is given by

$$\text{Proj}(S) = \bigcup_{J \subset I \text{ relevant}} \text{Spec}(R[\sigma_{J - I} \cap M]).$$

The (possibly nonseparated) union is taken with respect to the inclusions $J \subset J'$.

**Example 3.8.** Let $S = R[T_1, \ldots, T_k]$ be a polynomial algebra graded by $D = \mathbb{Z}$ so that all indeterminates have positive degree. Then Proj($S$) is the weighted projective space studied by Delorme [7], Mori [21], and Dolgachev [8].

**Example 3.9.** Here we construct a separated non-quasiprojective scheme defined by a single equation inside a multihomogeneous spectrum. Let

$$S = k[X_1, \ldots, X_4, Y_1, \ldots, Y_4, Z]$$

be a $\mathbb{Z}^2$-graded polynomial ring over a field $k$, with degrees $\text{deg}(X_i) = (1, 0)$, $\text{deg}(Y_j) = (0, 1)$, and $\text{deg}(Z) = (1, 1)$. Set $P = \text{Proj}(S)$ and consider the open subset $U = D_+(X_1Z) \cup D_+(Y_1Z)$. This is not separated: We have

$$\Gamma(D_+(X_1Z), \mathcal{O}_P) = k \left[ \frac{X_i}{X_1}, \frac{X_1Y_m}{Z} \right] \quad \text{and}$$

$$\Gamma(D_+(Y_1Z), \mathcal{O}_P) = k \left[ \frac{Y_j}{Y_1}, \frac{X_mY_1}{Z} \right],$$

(1)
with $1 \leq i,j,m \leq 4$. On the intersection $D_+(X_1Y_1Z) = D_+(X_1Z) \cap D_+(Y_1Z)$, these algebras generate the subalgebra

$$k \left[ \frac{X_i}{X_1}, \frac{Y_j}{Y_1}, \frac{X_1Y_1}{Z} \right] \subset k \left[ \frac{X_i}{X_1}, \frac{Y_j}{Y_1}, \frac{X_1Y_1}{Z}, \frac{Z}{X_1Y_1} \right] = \Gamma(D_+(X_1Y_1Z), \mathcal{O}_P),$$

which does not contain $Z/X_1Y_1$. To obtain separated subschemes, we have to kill $Z/X_1Y_1$. Consider the homogeneous polynomials of degree $(2,2)$

$$g = X_1Y_1Z + X_2^2Y_1^2 + X_1^2Y_2^2 \quad \text{and} \quad f = X_1Y_1Z + X_2^2Y_1^2 + X_1^2Y_2^2 + X_3X_4Y_3Y_4.$$  

Let $S' = S/(f)$, $P' = \text{Proj}(S')$ and $U' = U \cap P'$. Modulo $f$, the element $Z/X_1Y_1$ is generated by the algebras in (1), thus $U'$ is a separated scheme. It is, however, not quasiprojective. First observe that $S'$ is a factorial domain: $S'_{X_3X_4Y_3} \cong k[X_0, \ldots, X_4, Y_1, Y_2, Y_3, Z]_{X_3X_4Y_3}$ is factorial and $X_3, X_4, Y_3$ are prime in $S'$, because $g \in k[X_1, X_2, Y_1, Y_2, Z]$ is prime.

Choose points $x \in V_+(X_1, \ldots, X_4) \cap U'$ and $y \in V_+(Y_1, \ldots, Y_4) \cap U'$ (such points exist), and assume that they admit a common affine neighborhood $W \subset U'$. Then the preimage $V \subset \text{Spec}(S')$ is affine as well. By factoriality, $V = D(h)$ for some homogeneous $h \in S$ with $h \in (X_1Z, Y_1Z)$. Write $h = pX_1Z + qY_1Z$. Let $\deg(h) = (d_1, d_2)$ and suppose $d_1 \geq d_2$. Since $Y_1Z$ has degree $(1,2)$, it follows that $q \in (X_1, \ldots, X_4)$. But then $h(x) = 0$, contradiction.

4. Ample families and mappings to homogeneous spectra.

In this section, we shall relate homogeneous spectra to ample families. Let $X$ be a scheme, $D$ a finitely generated abelian group, and $\mathcal{B} = \bigoplus_{d \in D} \mathcal{B}_d$ a quasicoherent $D$-graded $\mathcal{O}_X$-algebra. We say that $\mathcal{B}$ is periodic if each stalk $\mathcal{B}_x$ is a periodic $\mathcal{O}_{X,x}$-algebra. For each homogeneous $f \in \Gamma(X, \mathcal{B})$, let $X_f \subset X$ be the largest open subset such that all multiplication maps $f^n : \mathcal{O}_X \to \mathcal{B}_n$ with $n \geq 0, d = \deg(f)$ are bijective.

**Proposition 4.1.** Let $S$ be a $D$-graded ring, $X = \text{Proj}(S)$ its homogeneous spectrum, $Y = \text{Spec}(S) - V(S_+) \text{ the relevant locus}$, and $\pi : Y \to X$ the natural projection. Then $\pi_*(\mathcal{O}_Y)$ is a periodic $\mathcal{O}_X$-algebra. Furthermore, $X_f = D_+(f)$ for each relevant $f \in S$.

**Proof.** By definition, $S_g$ is a periodic $S_{(g)}$-algebra for each relevant $g \in S$, so $\pi_*(\mathcal{O}_Y)$ is a periodic $\mathcal{O}_X$-algebra with $\pi_*(\mathcal{O}_Y)_0 = \mathcal{O}_X$. The inclusion $D_+(f) \subset X_f$ is obvious. To verify $X_f \subset D_+(f)$, it suffices to check that $f \in S_g$ is invertible for each relevant $g \in S$ with $D_+(g) \subset X_f$. Replacing $f$ by a positive multiple and $S_g$ by a suitable Veronese subring, the ring $S_g$ becomes isomorphic to the Laurent polynomial algebra $S_{(g)}[T_1^{\pm 1}, \ldots, T_r^{\pm 1}]$, and $f$ corresponds to a monomial $\lambda T^d$ with $\lambda \in S_{(g)}$. Hence $f$ is invertible. \(\Box\)
Next, we extend Grothendieck’s description ([11] Prop. 3.7.3) of mappings into homogeneous spectra to the multigraded case:

**Proposition 4.2.** Let $X$ be a scheme, $B$ a quasicoherent $D$-graded $\mathcal{O}_X$-algebra, $S$ a $D$-graded ring, and $\varphi : S \to \Gamma(X, B)$ a graded homomorphism. Set $U = \bigcup X_{\varphi(f)}$, where the union runs over all relevant $f \in S$. Then there is a natural morphism $r_{B,\varphi} : U \to \text{Proj}(S)$ and a commutative diagram

$$
\begin{array}{ccc}
U & \xrightarrow{r_{B,\varphi}} & \text{Spec}(B_U) \\
\downarrow \varphi & & \downarrow \varphi \\
\text{Proj}(S) & \xrightarrow{\varphi} & \text{Spec}(S) - V(S_+) \xrightarrow{\varphi} \text{Spec}(S).
\end{array}
$$

**Proof.** Each relevant $f \in S$ gives a homomorphism $S(f) \to \Gamma(X, B)(f)$, where we write $f$ instead of $\varphi(f)$. Furthermore, there is a homomorphism $\Gamma(X, B)(f) \to \Gamma(X_f, \mathcal{O}_X)$, $g/f^n \mapsto (f^n|X_f)^{-1}(g)$, where $(f^n|X_f)^{-1}$ is the inverse mapping for the bijective multiplication mapping $f^n|X_f : \mathcal{O}_X|X_f \to B_{nd}|X_f$. The composition defines a morphism $X_f \to D_+(f)$. You easily check that these morphisms coincide on the overlaps, and we obtain the desired morphism $r_{B,\varphi} : U \to \text{Proj}(S)$.

We write $r_{B,\varphi} : X \dashrightarrow \text{Proj}(S)$ for the morphism $r_{B,\varphi} : U \to \text{Proj}(S)$ and call it a rational map. Saying that a rational map is everywhere defined means $U = X$. In this case, we have a honest morphism $r_{B,\varphi} : X \to \text{Proj}(S)$.

**Corollary 4.3.** Let $S$ be a $D$-graded ring. For each morphism $r : X \to \text{Proj}(S)$, there is a quasicoherent periodic $D$-graded $\mathcal{O}_X$-algebra $B$ and a homomorphism $\varphi : S \to \Gamma(X, B)$ such that the rational map $r_{B,\varphi} : X \dashrightarrow \text{Proj}(S)$ is everywhere defined and coincides with $r : X \to \text{Proj}(S)$.

**Proof.** Let $Y = \text{Spec}(S) - V(S_+)$ be the irrelevant locus, $\pi : Y \to \text{Proj}(S)$ the canonical projection, and set $B = r^*(\pi_*\mathcal{O}_X))$.

We come to the characterization of ample families in terms of homogeneous spectra:

**Theorem 4.4.** Let $\mathcal{L}_1, \ldots, \mathcal{L}_r$ be a family of invertible sheaves on a quasicompact and quasiseparated scheme $X$. Then the following conditions are equivalent:

(i) The family $\mathcal{L}_1, \ldots, \mathcal{L}_r$ is ample.

(ii) The canonical rational map $X \dashrightarrow \text{Proj}(X, \bigoplus_{d \in \mathbb{N}} \mathcal{L}_d)$ is everywhere defined and an open embedding.

If $X$ is of finite type over a noetherian ring $R$, this is also equivalent with:
(iii) There is a finite family of sections \( f_i \in \Gamma(X, \mathcal{L}_d), i \in I \) and a \( D \)-graded polynomial algebra \( A = R[T_i]_{i \in I} \) such that the rational map \( X \dashrightarrow \text{Proj}(A) \) induced by \( T_i \mapsto f_i \) is everywhere defined and an embedding.

Proof. Set \( S = \Gamma(X, \bigoplus_{d \in \mathbb{N}^r} \mathcal{L}_d) \). We start with the implication (i) \( \Rightarrow \) (ii). According to Proposition 1.1, for each point \( x \in X \), there is a \( \mathbb{Q} \)-basis \( d_i \in \mathbb{N}^r \) and sections \( f_i \in \Gamma(X, \mathcal{L}_d) \) so that \( X_{f_i} \) are affine neighborhoods of \( x \). Consequently, \( f = f_1 \ldots f_r \in S \) is relevant, so the rational map \( X \dashrightarrow \text{Proj}(S) \) is everywhere defined. Fix a relevant \( f \in S \) so that \( X_f \) is affine. According to [10], the canonical map \( \Gamma(X, \bigoplus \mathcal{L}_d) \rightarrow \Gamma(X_f, \bigoplus \mathcal{L}_d) \) is bijective. Consequently, \( X_f \rightarrow D_+(f) \) is an isomorphism, so \( X \rightarrow \text{Proj}(S) \) is an open embedding. The reverse implication is trivial.

For the rest of the proof, suppose that \( X \) is of finite type over a noetherian ring \( R \). Assume that \( \mathcal{L}_1, \ldots, \mathcal{L}_r \) is ample. Choose finitely many relevant \( f_i \in S \) so that \( X_{f_i} \subset X \) form an affine open cover. Due to the assumption we may write \( \Gamma(X_{f_i}, \mathcal{O}_X) = R[h_{i_1}, \ldots, h_{i_m}] \). For suitable \( n \) we have \( f_{ij} := f^n h_{ij} \in S \). Let \( f_i \in S, i \in I \) be these elements all together.

Let \( R[T_i] \) be graded by \( d(T_i) = \deg(f_i) \) such that the natural mapping \( R[T_i] \rightarrow S \) is homogeneous. Then \( X_{f_i} \rightarrow D_+(f_i) \) are closed embeddings, since the ring morphisms are surjective, and so \( X \rightarrow \text{Proj}(A) \) is an embedding. Finally, the implication (iii) \( \Rightarrow \) (ii) is trivial. \( \square \)

Corollary 4.5. Let \( \mathcal{L}_1, \ldots, \mathcal{L}_r \) an ample family on \( X \). For each relevant element \( f \in \Gamma(X, \bigoplus_{d \in \mathbb{N}^r} \mathcal{L}_d) \), the open subset \( X_f \) is quasiaffine.

Proof. According to Theorem 4.4, we have an open embedding \( X_f \subset D_+(f) \). \( \square \)

Corollary 4.6. Let \( \mathcal{L}_1, \ldots, \mathcal{L}_r \) an ample family of invertible sheaves on \( X \). Set \( S = \Gamma(X, \bigoplus_{d \in \mathbb{N}^r} \mathcal{L}_d) \). Then the following conditions are equivalent:

(i) The open embedding \( r : X \rightarrow \text{Proj}(S) \) is an isomorphism.

(ii) For each relevant \( f \in S \), the quasiaffine open subset \( X_f \subset X \) is affine.

If the affine hull \( X \rightarrow X^\text{aff} \) is proper, this is also equivalent to:

(iii) The homogeneous spectrum \( \text{Proj}(S) \) is separated.

Proof. The equivalence (i) \( \iff \) (ii) follows from \( D_+(f) = X_f^\text{aff} \). Now assume that \( X \rightarrow X^\text{aff} \) is proper. The implication (i) \( \Rightarrow \) (iii) is trivial. To see the converse, we apply [11] Corollary 5.4.3 and infer that the open dense embedding \( X_f \rightarrow D_+(f) \) is proper, hence an isomorphism. \( \square \)

Finally, we generalize Hausen’s [17] characterization of divisorial varieties:

Corollary 4.7. Let \( X \) be a scheme of finite type over a noetherian ring \( R \). Then the following are equivalent:

(i) The scheme \( X \) is divisorial.
(ii) There is an embedding of $X$ into the homogeneous spectrum of a multi-graded $R$-algebra of finite type.

(iii) $X$ is embeddable into a simplicial torus embedding with affine diagonal.

Proof. If $X$ is divisorial, Theorem 4.4 ensures the existence of an embedding $X \subset \text{Proj}(S)$ with $S$ finitely generated. The implication (ii) $\Rightarrow$ (iii) follows from Proposition 3.4, and (iii) $\Rightarrow$ (i) is trivial. \qed

5. Cohomological characterization of ample families.

Throughout this section, $R$ is a noetherian ring and $X$ is a proper $R$-scheme. According to Serre’s Criterion ([12] Prop. 2.6.1), an invertible $\mathcal{O}_X$-module $L$ is ample if and only if for each coherent $\mathcal{O}_X$-module $F$ there is an integer $n_0$ so that $H^p(X, F \otimes L^n) = 0$ for all $p > 0$, $n > n_0$. The task now is to generalize this to ample families.

Let $\mathcal{A} = \bigoplus_{n \geq 0} A_n$ be a graded quasicoherent $\mathcal{O}_X$-algebra of finite type generated by $A_1$. Then $P = \text{Proj}(\mathcal{A})$ is a projective $X$-scheme and $\mathcal{O}_P(1)$ is an $X$-ample invertible sheaf. In general, however, $\mathcal{O}_P(1)$ is not ample in the absolute sense. More precisely:

**Proposition 5.1.** With the preceding notation, the invertible sheaf $\mathcal{O}_P(1)$ is ample if and only if for each coherent $\mathcal{O}_X$-module $F$, there is an integer $n_0$ so that $H^p(X, F \otimes A_n) = 0$ for $p > 0$ and $n > n_0$.

Proof. Let $h : P \rightarrow X$ be the canonical projection. First, suppose that $\mathcal{M} = \mathcal{O}_P(1)$ is ample. Choose $n_0 > 0$ so that the canonical map $A_n \rightarrow h_*(\mathcal{M}^n)$ is bijective and that $R^q h_*(h^*(\mathcal{F}) \otimes \mathcal{M}^n) = 0$ and $H^p(P, h^*(\mathcal{F}) \otimes \mathcal{M}^n) = 0$ holds for $p, q > 0$, $n > n_0$. Using the Leray–Serre spectral sequence we infer $H^p(X, h_*(h^*(\mathcal{F}) \otimes \mathcal{M}^n)) = 0$ for all $p > 0$, $n > n_0$. We claim that the adjunction map $F \otimes h_*(\mathcal{M}^n) \rightarrow h_*(h^*(\mathcal{F}) \otimes \mathcal{M}^n)$ is bijective for $n \gg 0$. Fix a point $x \in X$ and choose a finite presentation

\[ O_{x,x}^k \otimes h_*(\mathcal{M}^n) \rightarrow O_{x,x}^l \otimes h_*(\mathcal{M}^n) \rightarrow F_x \otimes h_*(\mathcal{M}^n) \rightarrow 0. \]

Then we have an exact sequence

\[ O_{x,x}^k \otimes h_*(\mathcal{M}^n) \rightarrow O_{x,x}^l \otimes h_*(\mathcal{M}^n) \rightarrow F_x \otimes h_*(\mathcal{M}^n) \rightarrow 0, \]

and another exact sequence

\[ h_*(O_{P}^k \otimes \mathcal{M}^n) \rightarrow h_*(O_{P}^l \otimes \mathcal{M}^n) \rightarrow h_*(h^*(\mathcal{F}) \otimes \mathcal{M}^n) \rightarrow R^1 h_*(G \otimes \mathcal{M}^n) \]

on $\text{Spec}(O_{x,x})$, where $G$ is the kernel of $O_{P}^k \rightarrow h^*(\mathcal{F})$. But $R^1 h_*(G \otimes \mathcal{M}^n) = 0$ for $n \gg 0$. By the 5-Lemma, the Claim is true locally around $x$. Using quasicompactness, we infer that the Claim holds globally. Enlarging $n_0$ if necessary, we have $H^p(X, F \otimes A_n) = 0$ for $p > 0$ and $n > n_0$ as desired. The converse is similar. \qed
Let $\mathcal{L}$ be an invertible $\mathcal{O}_X$-module. The idea now is to consider coherent submodules $\mathcal{K} \subset \mathcal{L}$. Note that such submodules correspond to quasi-coherent graded subalgebras $\bigoplus_{n \geq 0} \mathcal{K}^n \subset \mathcal{S}(\mathcal{L})$ locally generated by terms of degree one. In turn, we obtain a projective $X$-scheme $P = \text{Proj}(\bigoplus_{n \geq 0} \mathcal{K}^n)$ endowed with an $X$-ample invertible sheaf $\mathcal{O}_P(1)$. Locally, $P \to X$ looks like the blowing-up of an ideal.

**Proposition 5.2.** Let $\mathcal{L}$ be an invertible $\mathcal{O}_X$-module, and $x \in X$ a closed point. Then the following are equivalent:

(i) For some $n > 0$, there is a section $f \in H^0(X, \mathcal{L}^n)$ so that $X_f$ is an affine neighborhood of $x$.

(ii) For some $d > 0$, there is a coherent subbundle $\mathcal{K} \subset \mathcal{L}^d$ with $\mathcal{K}_x \subset \mathcal{L}_x^d$ bijective, so that the graded $\mathcal{O}_X$-algebra $\mathcal{A} = \bigoplus_{n \geq 0} \mathcal{K}^n$ satisfies the equivalent conditions of Proposition 5.1.

**Proof.** First, we check (i) $\Rightarrow$ (ii). Set $U = X_f$. According to [19] Proposition 5.4, there is a blowing-up $h : P \to X$ with center disjoint from $U$, together with an effective ample Cartier divisor $D \subset P$ satisfying $\text{Supp}(D) = P - U$. Set $\mathcal{M} = \mathcal{O}_P(D)$, such that $P = \text{Proj}(\bigoplus_{n \geq 0} h_*(\mathcal{M}^n))$. Replacing $D$ by a suitable multiple, we may assume that $\bigoplus_{n \geq 0} h_*(\mathcal{M}^n)$ is generated by terms of degree one.

The identity $g_U : h_*(\mathcal{M})|U \to \mathcal{O}_X|U$ extends to a mapping $g : h_*(\mathcal{M}) \to \mathcal{L}^d$ for $d \gg 0$. This map is injective because a nonzero section
t $t \in \Gamma(P, \mathcal{O}_P(D)) = \Gamma(X, h_*(\mathcal{M}))$
do not vanish on $U$. Let $\mathcal{K} \subset \mathcal{L}^d$ be the image of $g : h_*(\mathcal{M}) \to \mathcal{L}^d$. Then $\mathcal{K}^n = h_*(\mathcal{M}^n)$, because $\bigoplus_{n \geq 0} h_*(\mathcal{M}^n)$ is generated by terms of degree one.

On $P = \text{Proj}(\bigoplus \mathcal{K}^n)$ we have $\mathcal{O}_P(1) = \mathcal{M}$, which is ample as desired.

Now we check (ii) $\Rightarrow$ (i). Set $P = \text{Proj}(\mathcal{A})$ and let $h : P \to X$ be the canonical morphism. Choose an affine open neighborhood $U \subset X$ of $x$ so that the induced projection $h^{-1}(U) \to U$ is an isomorphism. Since $\mathcal{O}_P(1)$ is ample, there is an integer $m > 0$ and a section $g \in H^0(X, \mathcal{K}^m)$ so that the induced section $g' \in H^0(P, \mathcal{O}_P(m))$ vanishes on $P - h^{-1}(U)$ and is nonzero on the point $h^{-1}(x)$. Let $f$ be the image of $g$ under the inclusion $H^0(X, \mathcal{K}^m) \hookrightarrow H^0(X, \mathcal{L}^d)$. Let $y \in X_f$. Then $\mathcal{K}^m_y = \mathcal{L}^{dm}_y$, and $X_f \cong X \subset U$ and $X_f$ is affine.

The preceding result leads to a characterization of ample families in terms of cohomology:

**Theorem 5.3.** Let $R$ be a noetherian ring and $X$ a proper $R$-scheme. A family $\mathcal{L}_1, \ldots, \mathcal{L}_r$ of invertible $\mathcal{O}_X$-modules is ample if and only if for each closed point $x \in X$, there is $d \in \mathbb{N}^r$ and a coherent subsheaf $\mathcal{K} \subset \mathcal{L}^d$ with
AMPLE FAMILIES AND HOMOGENEOUS SPECTRA 227

\[ K_x = \mathcal{L}^d \] so that for each coherent \( \mathcal{O}_X \)-module \( F \) there is an integer \( n_0 \) with \( H^p(X, F \otimes K^n) = 0 \) for all \( p > 0, n > n_0 \).

**Proof.** This follows directly from Proposition 5.2. \( \square \)

6. Algebraization of formal schemes via ample families.

Throughout this section, \((R, \mathfrak{m}, k)\) denotes a complete local noetherian ring, and \( \mathfrak{X} \rightarrow \text{Spf}(R) \) is a proper formal scheme. Such formal schemes frequently occur as formal solutions for problems related to moduli spaces and deformation theories. A natural question is whether such a formal scheme is algebraizable. This means that there is a proper scheme \( \mathfrak{X} \rightarrow \text{Spec}(R) \) whose \( \mathfrak{m} \)-adique completion is isomorphic to \( \mathfrak{X} \). Grothendieck's Algebraization Theorem ([12] Thm. 5.4.5) asserts that \( \mathfrak{X} \) is algebraizable if there is an invertible \( \mathcal{O}_X \)-module whose restriction to the closed fiber \( \mathfrak{X}_0 = \mathfrak{X} \otimes k \) is ample. Here is a generalization:

**Theorem 6.1.** Let \( \mathfrak{X} \) be a proper formal scheme as above, and \( \mathcal{L}_1, \ldots, \mathcal{L}_r \) a family of invertible \( \mathcal{O}_X \)-modules. Suppose that for each closed point \( x \in \mathfrak{X} \), there is a degree \( d \in \mathbb{N}^r \) and a coherent submodule \( \mathcal{K} \subset \mathcal{L}_x^d \) with \( \mathcal{K}_x \subset \mathcal{L}_x^d \) bijective, so that \( \mathcal{O}_{P_0}(1) \) is ample on \( P_0 = \text{Proj}(\oplus_{m \geq 0} \mathcal{K}^m / I \mathcal{K}^m) \). Then the formal scheme \( \mathfrak{X} \) is algebraizable.

**Proof.** Set \( \mathfrak{X}_n = \mathfrak{X} \otimes R / \mathfrak{m}^{n+1} \), such that \( \mathfrak{X} = \lim_{\rightarrow} \mathfrak{X}_n \), and let \( I \subset \mathcal{O}_X \) be the ideal of the closed fiber \( \mathfrak{X}_0 \subset \mathfrak{X} \). As in the Proof of Proposition 5.2, there is an integer \( m > 0 \) and a global section \( s_0 \in H^0(\mathfrak{X}_0, \mathcal{K}_m / I \mathcal{K}_m) \) so that the induced section \( t_0 \in \Gamma(\mathfrak{X}_0, \mathcal{L}_m / I \mathcal{L}_m) \) defines an affine open neighborhood \( x \in (\mathfrak{X}_0)_0 \). We seek to extend such sections to formal sections.

For each \( m \geq 0 \), set \( A_m = \bigoplus_{n \geq 0} I^n \mathcal{K}_m / I^{n+1} \mathcal{K}_m \), and let \( A = \bigoplus_{m \geq 0} A_m \) be the corresponding \( \mathbb{N} \)-graded quasicoherent \( \mathcal{O}_{X_0} \)-algebra. Consider its homogeneous spectrum \( P = \text{Proj}(A) \). We claim that the invertible sheaf \( \mathcal{O}_P(1) \) is ample. To see this, let \( \mathfrak{X}' \) be the affine \( X_0 \)-scheme defined by \( \mathcal{O}_{\mathfrak{X}'} = \bigoplus_{n \geq 0} \mathcal{L}_n / \mathcal{L}_{n+1} \). Note that \( \mathfrak{X}' \) is proper over the noetherian ring \( R' = \bigoplus \mathfrak{m}^n / \mathfrak{m}^{n+1} \). Set

\[ P' = \text{Proj} \left( \bigoplus_{m \geq 0} \left( \bigoplus_{n \geq 0} \mathcal{L}_n / I^m \mathcal{L}_n \otimes \mathcal{K}_m / I \mathcal{K}_m \right) \right), \]

where the homogeneous spectrum is taken with respect to the grading \( m \in \mathbb{Z} \). We have \( P' = \mathfrak{X}' \times_{X_0} P_0 \) and conclude that \( \mathcal{O}_{P'}(1) \) is ample.

The surjective mapping \( \mathcal{L}_n \otimes \mathcal{K}_m / I \mathcal{K}_m \to \mathcal{L}_n \mathcal{K}_m / I \mathcal{K}_m \) induces a surjective mapping

\[
\begin{align*}
\mathcal{L}_n / I^{n+1} \otimes_{X_0} \mathcal{K}_m / I \mathcal{K}_m &\to \mathcal{L}_n \otimes_{X_0} \mathcal{K}_m \otimes_{X_0} \mathcal{O}_{X_0} \\
&\quad \to \mathcal{L}_n \mathcal{K}_m / I \mathcal{K}_m.
\end{align*}
\]
This yields a closed embedding $P \subset P'$, showing that $O_P(1)$ is ample. To proceed, consider the exact sequence

$$H^0(X_n, \mathcal{K}^m/I^{n+1}\mathcal{K}^m) \longrightarrow H^0(X_{n-1}, \mathcal{K}^m/I^n\mathcal{K}^m) \longrightarrow H^1(X_0, I^n\mathcal{K}^m/I^{n+1}\mathcal{K}^m).$$

By Proposition 5.1, there is an integer $m_0 > 0$ so that the group on the right is zero for all $m \geq m_0$ and all $n \geq 0$. Passing to a suitable multiple if necessary, we can lift our section $s_0 \in H^0(X_0, \mathcal{K}^m/I^m\mathcal{K}^m)$ to a formal section $s \in H^0(X, \mathcal{K}^m)$. Therefore, the section $t_0 \in \Gamma(X_0, \mathcal{L}_{dm}/I\mathcal{L}_{dm})$ lifts to a formal section $t \in \Gamma(X, \mathcal{L}_{dm})$.

Using such formal sections, you construct as in the Proof of Theorem 4.4 a finitely generated $\mathbb{N}^r$-graded polynomial $R$-algebra $S$ and a compatible sequence of embeddings $X_n \subset \text{Proj}(S)$. Choose an open neighborhood $U \subset \text{Proj}(S)$ so that $X_n \subset U$ are closed embeddings. Let $I_n \subset S$ be the graded ideal of the closed embedding $X_n \subset \text{Proj}(S)$, and set $I = \bigcap_{n \geq 0} I_n$. Then $X = \text{Proj}(S/I) \cap U$ is the desired algebraization of the formal scheme $X$.

**Remark 6.2.** If we have $\mathcal{K} = \mathcal{L}^d$, then $P_0 = X_0$, such that the formal sheaf $\mathcal{L}$ is ample on the closed fiber $X_0$. In this case, Grothendieck’s Algebraization Theorem ensures that $X$ is algebraizable.

**Question 6.3.** The assumptions in Theorem 6.1 imply that the restriction of the family $\mathcal{L}_1, \ldots, \mathcal{L}_r$ to the closed fiber is ample. A natural question to ask: Given a proper formal scheme with a family of invertible sheaves whose restriction to the closed fiber is ample – is the formal scheme algebraizable?

**Acknowledgement.** We thank Professor Uwe Storch for helpful suggestions. The second author is grateful to the M.I.T. Mathematical Department for its hospitality, and thanks the DFG for financial support.

**References**


Received April 13, 2001.

Mathematische Fakultät
Ruhr-Universität
44780 Bochum
Germany
E-mail address: brenner@cobra.ruhr-uni-bochum.de

Mathematische Fakultät
Ruhr-Universität
44780 Bochum
Germany
E-mail address: s.schroeer@ruhr-uni-bochum.de

M.I.T. Department of Mathematics
Cambridge, MA 02139-4307
A STABILITY CRITERION FOR EXTREMALS OF ELLIPTIC PARAMETRIC FUNCTIONALS

ULRICH CLARENZ

In this paper we consider parametric integrals with elliptic integrands depending on the surface normal. The main result is a stability criterion for extremal immersions of those functionals, containing a result of Barbosa and do Carmo for minimal surfaces as a special case. Using similar techniques we are also able to show a condition for instability. The last section contains a simple proof of the fact that the surface normal of extremals of parametric integrals is topologically equivalent to a holomorphic function.

1. Introduction.

Since the last century, mathematicians have tried to find criteria for stability and instability of a given minimal surface. First important criteria of stability and instability were proven by Schwarz (for a survey of his results concerning stability of minimal surfaces see [9, pp. 90-116]).

He showed that a minimal surface is stable if its spherical image is fully contained in a half-sphere [9, p. 99].

In 1976, Barbosa and do Carmo were able to generalize this criterion considerably. They proved that if the spherical image of a given minimal surface has area less than $2\pi$, the surface is stable [1].

Barbosa and do Carmo generalized this result in several directions as e.g., for different target manifolds and for general dimensions [2].

Ruchert proved a similar criterion for the case of surfaces of constant mean curvature [11].

In [7], Fischer-Colbrie and Schoen were able to give a shorter proof of the result of Barbosa and do Carmo [1]. Here we will follow the idea of Fischer-Colbrie and Schoen to prove a more general theorem for critical points of parametric functionals. In our context a parametric functional is defined for an immersion $X : M \to \mathbb{R}^3$ with normal mapping $N : M \to S^2$ by

$$\mathcal{F}(X) := \int_M F(N) \, dA,$$

where $M$ is a two-dimensional, orientable manifold.
The integrand \( F : \mathbb{R}^3 \to \mathbb{R} \) is a 1-homogeneous function, i.e., \( F(tz) = tF(z) \) for all \( t > 0 \). Immersions satisfying the Euler equation of \( F \) are called \( F \)-extremals or \( F \)-minimal surfaces.

Furthermore we call \( D \subset M \) a stable domain of an \( F \)-extremal \( X \) if for all \( \varphi \in C_0^\infty(D) \) with \( \varphi \not\equiv 0 \) the second variation

\[
\delta^2 F(X, \varphi) = \frac{d^2}{d\epsilon^2} F(X_\epsilon)_{|\epsilon=0}, \quad X_\epsilon = X + \epsilon \varphi N,
\]

is positive (Definition 4.1).

In this paper we only deal with \textit{elliptic integrands}. Ellipticity means that

\[ F_{zz}(z) : z^\perp \to z^\perp \]

is a positive definite endomorphism of \( z^\perp \) for all \( z \in \mathbb{R}^3 - \{0\} \).

Here we state our main result:

\textbf{Theorem 1.1.} Let \( X : M \to \mathbb{R}^3 \) be an \( F \)-extremal immersion of an elliptic integrand \( F \). If the area of the spherical image of a smooth domain \( D \subset M \) is smaller than a positive number \( a_F \) depending on \( F \), then \( D \) is a stable domain of \( X \).

In the case of \( F(z) = |z| \), the number \( a_F \) is \( 2\pi \). Furthermore, \( a_F \) depends continuously on \( F \) w.r.t. \( C^0(S^2) \).

Using similar techniques as in the proof of Theorem 1.1 we can also prove Theorem 4.5, a generalization for \( F \)-extremals of a Schwarz criterion for instability (see [1, Theorem 2.7]).

The paper is organized as follows:

In Section 2 we define the notion of a \textit{degenerate metric} \( \tilde{g} \) on a Riemannian manifold \( (M, g) \) and give important properties of them. Degenerate metrics \( \tilde{g} \) are allowed to have isolated singularities and \( g \) and \( \tilde{g} \) are related by the equivalence given in Definition 2.1. One example of such a degenerate metric is \( g_N(V, W) := g(DN(V), DN(W)) \) for a minimal immersion \( X \) with canonical metric \( g \) and normal mapping \( N \) (in fact \( g_N = -Kg \), where \( K \) is the Gauß-curvature).

In Section 3 we will study basic properties of parametric functionals. Especially we will see that for \( F \)-extremals, \( g_N \) generally is not conformal as in the minimal surface case but still degenerate in the sense of Definition 2.1.

In Section 4 we prove the main result Theorem 1.1. First we give an estimate of the second variation \( \delta^2 F \), showing a connection between stability and eigenvalue problems for the Laplacian \( \Delta_{S^2} \) on the sphere \( S^2 \) (Proposition 4.2):

\[
\delta^2 F(X, \varphi) \geq C_F \int_M |\text{grad}_N \varphi|^2_N - 2c_F \varphi^2 dA_N.
\]

Here \( c_F \geq 1 \) and \( C_F \) are constants depending on \( F \). The subscript \( N \) indicates the relation to \( g_N \). Now we use a Faber-Krahn argument to show
that the right-hand side of (1) is positive. To this aim we take a first (positive) eigenfunction $u$ of $\Delta_{S^2}$ of the spherical image of a domain $D$, i.e.,

$$\Delta_{S^2} u + \mu_1 u = 0.$$

Lifting this equation onto $M$ we obtain a positive function $v$ on $M$ (because $N$ is an open mapping) satisfying outside the singularities of $g_N$:

$$(2) \quad \Delta_N v + \mu_1 v = 0.$$

At this point it is crucial that $g_N$ is a degenerate metric. Using Proposition 2.4, a generalization of [7, Corollary 1], Equation (2) implies the inequality $\int |\text{grad}_N \varphi|^2_N - \mu_1 \varphi^2 dA_N > 0$ for all $\varphi \in C^\infty_0(D)$ and $\varphi \not\equiv 0$. Now we can choose $a_F$ dependent on $c_F$ such that $\mu_1 > 2c_F$.

In the above reasoning we have used that $N$ is an open mapping. This was proven by Sauvigny [12, p. 94]. As a by-product of our considerations we can give a simple proof of the fact that for $F$-extremals $N$ is a local branched covering. This is the content of Section 5.

2. Degenerate metrics.

Let $(M, g)$ be a two-dimensional Riemannian manifold. The key notion for our considerations is given in the following:

**Definition 2.1.** A symmetric bilinear form $\tilde{g}$ on $(M, g)$ is a degenerate metric if for all $V \in T_p M$

$$\Gamma h_1 g(V, V) \leq \tilde{g}(V, V) \leq \Gamma h_2 g(V, V),$$

where $h_1, h_2 > 0, \Gamma \geq 0$ are smooth functions on $M$. Furthermore, the set $\{ \Gamma = 0 \}$ consists of isolated points only.

**Lemma 2.2.** Let $g_1, g_2$ be two metrics on a manifold $M$. If for all $V \in T_p M$

$$(3) \quad g_1(V, V) \leq g_2(V, V)$$

then we have for all smooth functions $\varphi$:

$$|\text{grad}_1 \varphi|_1 \geq |\text{grad}_2 \varphi|_2,$$

where the index indicates the corresponding metric.

**Proof.** We have the identity

$$g_1(\text{grad}_1 \varphi, V) = g_2(\text{grad}_2 \varphi, V),$$

for all $V \in T_p M$. Using the Cauchy-Schwarz inequality, this leads to

$$|g_2(\text{grad}_2 \varphi, V)| \leq |\text{grad}_1 \varphi|_1 |V|_1 \leq |\text{grad}_1 \varphi|_1 |V|_2,$$

because of inequality (3). Setting $V = \text{grad}_2 \varphi$ gives the assertion. \square

As a direct consequence of the above lemma we obtain:
Corollary 2.3. If \( \tilde{g} \) is a degenerate metric on a Riemannian manifold \((M, g)\), the following inequality holds on \( \{ \Gamma \neq 0 \} \):
\[
\Gamma h_1 |\text{grad}_g \varphi|_g^2 \leq |\text{grad}_g \varphi|_g^2 \leq \Gamma h_2 |\text{grad}_g \varphi|_g^2.
\]

The next proposition is a generalization of [7, Corollary 1]:

Proposition 2.4. Let \((M, g)\) be a Riemannian manifold with degenerate metric \( \tilde{g} \). Furthermore let \( q \) be a smooth function on \( M \). Assume that there is a smooth function \( v \) on \( M \), \( v > 0 \) on a domain \( D \), with
\[
\Delta_g v - q v = 0 \tag{4}
\]
on \( \{ \Gamma > 0 \} \cap D \). Then for all \( \varphi \in C_0^\infty(D) \) we have
\[
\int_D |\text{grad}_g \varphi|_g^2 + q \varphi^2 \, dA_g \geq 0.
\]

Note that one can estimate the form \( |\text{grad}_g \varphi|_g^2 \, dA_g \) by \( |\text{grad}_g \varphi|_g^2 h_2 / h_1 \, dA_g \). Therefore the integral \( \int_D |\text{grad}_g \varphi|_g^2 \, dA_g \) is well-defined.

Proof of Proposition 2.4. The zeroes of \( \Gamma \) are isolated, i.e., \( \text{supp} \, \varphi \cap \overline{D} \cap \{ \Gamma > 0 \} \) is a finite set \( \{ p_1, \ldots, p_k \} \). In the following reasoning we may assume \( p_i \notin \partial D \). Let \( B_\epsilon(p_i) \) be a geodesic ball in \( M \) (measured in the \( g \)-metric) of radius \( 0 < \epsilon \ll 1 \), such that
\[
V_\epsilon := \bigcup_{i=1}^k B_\epsilon(p_i)
\]
is a disjoint union. Because of (4) we have with \( w = \log v \)
\[
\Delta_{\tilde{g}} w = q - |\text{grad}_{\tilde{g}} w|_{\tilde{g}}^2.
\]
Integration by parts leads to
\[
\int_{D - V_\epsilon} \varphi^2 |\text{grad}_{\tilde{g}} w|_{\tilde{g}}^2 - q \varphi^2 \, dA_{\tilde{g}}
\]
\[
= - \int_{D - V_\epsilon} \varphi^2 \Delta_{\tilde{g}} w \, dA_{\tilde{g}}
\]
\[
= \int_{D - V_\epsilon} 2 \varphi \tilde{g} (\text{grad}_{\tilde{g}} \varphi, \text{grad}_{\tilde{g}} w) \, dA_{\tilde{g}}
\]
\[
+ \sum_{i=1}^k \int_{\partial B_\epsilon(p_i)} \varphi^2 \tilde{g} (\text{grad}_{\tilde{g}} w, \nu) \, d\sigma_{\tilde{g}},
\]
with $\varphi \in C^\infty_0(D)$ and $\tilde{\nu}$ being a suitable unit normal on $\partial B_\epsilon(p_i)$. Using the fact that $\tilde{g}$ is a degenerate metric on $(M,g)$, we obtain the following estimate:

$$
\left| \int_{\partial B_\epsilon(p_i)} \varphi^2 \tilde{g}(\text{grad}_\tilde{g} w, \tilde{\nu}) \, d\sigma_{\tilde{g}} \right|
\leq \int_{\partial B_\epsilon(p_i)} \varphi^2 |\text{grad}_\tilde{g} w|_{\tilde{g}} \, d\sigma_{\tilde{g}}
\leq \int_{\partial B_\epsilon(p_i)} \varphi^2 |\text{grad}_g w|_g \sqrt{h_2/h_1} \, d\sigma_g = O(\epsilon), \quad \text{with } \epsilon \to 0.
$$

Thus we have

$$
\int_{D - V} \varphi^2 |\text{grad}_g w|_g^2 - q\varphi^2 \, dA_{\tilde{g}} = \int_{D - V} 2\varphi \tilde{g}(\text{grad}_\tilde{g} w, \text{grad}_\tilde{g} \varphi) \, dA_{\tilde{g}} + O(\epsilon).
$$

This implies

$$
O(\epsilon) \leq \int_{D - V} |\text{grad}_\tilde{g} \varphi|_{\tilde{g}}^2 + q\varphi^2 \, dA_{\tilde{g}}.
$$

3. Parametric functionals.

Now we consider immersed surfaces $X : M \to \mathbb{R}^3$, where $M$ is a two-dimensional and oriented manifold, equipped with the metric $g(V,W) = \langle DX(V), DX(W) \rangle$ for $V, W \in T_pM$ and a normal mapping

$$
N : M \to S^2.
$$

A parametric functional $\mathcal{F}$ is given by a smooth 1-homogeneous integrand

$$
F : \mathbb{R}^3 \to \mathbb{R}.
$$

The corresponding functional $\mathcal{F}$ is defined by $\mathcal{F}(X) := \int_M F(N) \, dA$.

As a version of $F_{zz}$ on $TM$ we define

$$
A_F : T_pM \to T_pM
V \mapsto DX^{-1} F_{zz}(N(p)) DX(V).
$$

**Definition 3.1.** An integrand $F$ is called **elliptic** if the linear mapping

$$
F_{zz}(z) : z^\perp \to z^\perp
$$

is positive definite for all $z \in \mathbb{R}^3 - \{0\}$.

Using the endomorphism $A_F$ the Euler equation of $\mathcal{F}$ is: $-\text{tr} (A_F \circ S) = 0$, where $S$ is the shape operator $DX \circ S := DN$ (for the corresponding computation see, e.g., [4] or [5]). The trace $H_F = -\text{tr} (A_F \circ S)$ is called the $F$-mean curvature $H_F$ generalizing the classical mean curvature $H = -\text{tr} S$. 

□
Now we want to prove for extremals $X$ of $\mathcal{F}$ that the bilinear form
\[ g_N(V,W) := \langle DN(V), DN(W) \rangle, \quad V,W \in T_pM \]
is a degenerate metric on $(M,g)$.

First, we apply the Cayley-Hamilton Theorem and get
\[ (A_F S)^2 + K_F \text{id}_{T_pM} = 0, \quad (6) \]
where $K_F := \det(A_F \circ S) = (\det A_F) K$. The so called $F$-Gauß curvature $K_F$ generalizes the classical Gauß curvature $K = \det S$. Equation (6) implies for all $V \in T_pM$ the identity
\[ g(A_F SV, SV) = -K_F g(A_F^{-1} V, V). \quad (7) \]

In the following, the eigenvalues of $A_F$ are denoted by $\lambda_1 \leq \lambda_2$. Now we estimate the left- and right-hand side of (7) for elliptic integrands.

\[
\begin{align*}
\lambda_1 g_N(V,V) & \leq g(A_F SV, SV) \leq \lambda_2 g_N(V,V), \\
-K_F \frac{1}{\lambda_2} g(V,V) & \leq -K_F g(A_F^{-1} V, V) \leq -K_F \frac{1}{\lambda_1} g(V,V).
\end{align*}
\]

Because of (7) we arrive at
\[ -K_F \frac{1}{\lambda_2} g(V,V) \leq g_N(V,V) \leq -K_F \frac{1}{\lambda_1} g(V,V) \]
and keeping in mind $K_F = \lambda_1 \lambda_2 K$ we state:

**Proposition 3.2.** Let $F$ be an elliptic integrand. For an $\mathcal{F}$-minimal surface $X : M \to \mathbb{R}^3$ the bilinear form $g_N$ is a degenerate metric on $(M,g)$. More precisely we have
\[ -K \frac{\lambda_1}{\lambda_2} g(V,V) \leq g_N(V,V) \leq -K \frac{\lambda_2}{\lambda_1} g(V,V). \]

For the proof of the above proposition it remains to show that $\{K = 0\}$ is a set of isolated points but this is a result of Sauvigny [13, p. 53].

\[ \text{4. Stability.} \]

In this part we assume that $F$ is an elliptic integrand. First we give a definition of stability.

**Definition 4.1.** Let $X : M \to \mathbb{R}^3$ be an $\mathcal{F}$-extremal immersion. The surface $X$ is called stable if for all $\varphi \in C^\infty_0(M)$ with $\varphi \not\equiv 0$ the relation
\[ \int_M g(A_F \text{grad}\varphi, \text{grad}\varphi) + \text{tr} A_F K \varphi^2 dA > 0, \quad (8) \]
is satisfied, where $K$ is the Gauß-curvature of $X$. We say that $D \subset M$ is a stable domain if (8) is fulfilled for all $\varphi \in C^\infty_0(D)$ with $\varphi \not\equiv 0$.
This definition generalizes the notion of stable minimal surfaces (see e.g., [6, p. 84] or [9, p. 96]; note that in this case $A_F = \text{id}_{T_pM}$ and $\text{tr} A_F = 2$). Its motivation is as follows:

If we consider $\varphi \in C_0^\infty(M)$ and the related disturbed surface $X_\epsilon = X + \epsilon \varphi N$, then the second variation $\delta^2 F(X, \varphi) = \frac{d^2}{d\epsilon^2} F(X_\epsilon)_{\epsilon=0}$ in direction $\varphi$ is given by the quadratic form in (8). The corresponding computation can be found in [10] or in [13].

In the following, all notions with subscript $N$ are related to the degenerate metric $g_N$. For the proof of the main result, we start with:

**Proposition 4.2.** Let $X : M \to \mathbb{R}^3$ be an $F$-critical immersion. Then we can conclude that $D \subset M$ is a stable domain if for all $\varphi \in C_0^\infty(D)$ and $\varphi \not\equiv 0$

$$\int_M |\text{grad}_N \varphi|^2 - 2c_F \varphi^2 dA_N > 0,$$

where $c_F := \max_{S^2} \left( \frac{\lambda_1 + \lambda_2}{2} \right) / \min_{S^2} \left( \frac{\lambda_1^2}{\lambda_2^2} \right)$.

**Proof.** Using the ellipticity of $F$ and the fact that $g_N$ is a degenerate metric we obtain because of Lemma 2.2 and Proposition 3.2

$$|\text{grad} \varphi|^2 \geq -K \frac{\lambda_1}{\lambda_2} |\text{grad}_N \varphi|^2_N.$$

This leads to

$$\delta^2 F(X, \varphi) = \int_M g(A_F \text{grad} \varphi, \text{grad} \varphi) + (\lambda_1 + \lambda_2) K \varphi^2 dA$$

$$\geq \int_M \lambda_1 |\text{grad} \varphi|^2 + (\lambda_1 + \lambda_2) K \varphi^2 dA$$

$$\geq \int_M -K \frac{\lambda_1^2}{\lambda_2} |\text{grad}_N \varphi|^2_N + (\lambda_1 + \lambda_2) K \varphi^2 dA.$$

The equation $\text{tr}(A_F S) = 0$ characterizing $F$-extremals implies the inequality $\det(A_F S) \leq 0$ and the ellipticity of $F$ gives $K = \det S \leq 0$. Therefore we have $dA_N = -K dA$. This completes the proof because of

$$\delta^2 F(X, \varphi) \geq \int_M \frac{\lambda_1^2}{\lambda_2} |\text{grad}_N \varphi|^2_N - (\lambda_1 + \lambda_2) \varphi^2 dA_N,$$

$$\geq \min_{S^2} \left( \frac{\lambda_1^2}{\lambda_2} \right) \int_M |\text{grad}_N \varphi|^2_N - 2c_F \varphi^2 dA_N.$$

The above proposition shows a connection between the stability of $F$-extremals and eigenvalue problems for the Laplacian on the sphere $S^2$. 

□
We will denote the first eigenvalue w.r.t. the Laplacian of a domain $D \subset S^2$ by $\mu_1(D)$. For a proof of the following proposition, we refer to [2, pp. 19, 20] and [3, pp. 50]:

**Proposition 4.3.** Assume that $D \subset S^2$ is a domain of area $A(D) = A$. Then we have:

(i) $A \leq 2\pi$ implies $\mu_1(D) \geq \frac{4\pi}{A}$,

(ii) $A \geq 2\pi$ implies $\mu_1(D) \geq \frac{4\pi - A}{A}$.

If in addition $D$ is a geodesic disc on $S^2$, then

(iii) $\mu_1(D) \leq \frac{4\pi}{A}$ for $A \geq 2\pi$,

(iv) $\lim_{A \to 4\pi} \mu_1(D) = 0$.

We see that for all $\mu \geq 2$ there is a spherical cap in $S^2$ whose first eigenvalue of the Laplacian is exactly $\mu$. The area of such a spherical cap is denoted by $a(\mu)$. For elliptic integrands we define $a_F := a(2c_F)$.

This enables us to prove the main result Theorem 1.1.

**Proof of Theorem 1.1.** Let $\Delta_{S^2}$ be the Laplacian on $S^2$. On $\{ K \neq 0 \}$ the normal $N$ is a local isometry between $S^2$ and $(M, g_N)$. We assume that $\mu_1$ is the first eigenvalue of $N(D)$ on $S^2$. Then there is a positive function $u$ in the interior of $N(D)$ satisfying $u|_{\partial N(D)} = 0$ and

$$\Delta_{S^2} u + \mu_1 u = 0.$$ 

In [13] it is proven that $N$ is an open mapping (see also Section 5). Therefore we can conclude that $v := u \circ N$ is positive in $D$ and satisfies on $K \neq 0$

$$\Delta_N v + \mu_1 v = 0.$$ 

Because of Proposition 2.4 we have for all $\varphi \in C_0^\infty(D)$

$$\int_D |\text{grad}_N \varphi|_N^2 - \mu_1 \varphi^2 dA_N \geq 0.$$ 

It is a well-known fact that the spherical caps on $S^2$ are minimizers of the first eigenvalue of $\Delta_{S^2}$ among all domains in $S^2$ of the same area [2, p. 18] and therefore $\mu_1 > 2c_F$. Thus we can conclude:

$$\int_D |\text{grad}_N \varphi|_N^2 - 2c_F \varphi^2 dA_N$$

$$= \int_D |\text{grad}_N \varphi|_N^2 - \mu_1 \varphi^2 dA_N + (\mu_1 - 2c_F) \int_D \varphi^2 dA_N > 0.$$
Using Proposition 4.2, stability is proven.

In case of the area-functional, i.e., \( F(z) = |z| \), the constant \( c_F \) is exactly 1. Therefore, for this functional we have \( a_F = a(2) = 2\pi \) and Theorem 1.1 contains the main result of [1] as a special case. □

The proof of the following proposition is very similar to the proof of Proposition 4.2:

**Proposition 4.4.** Assume that \( X : M \to \mathbb{R}^3 \) is an \( F \)-extremal and that there is a \( \varphi \in C^{\infty}(D) \cap C^0(\overline{D}), D \subset S^2 \), satisfying \( \varphi_{\partial D} = 0 \) and

\[
\int_D |\nabla_N \varphi|^2_N - 2d_F \varphi^2 \, dA_N < 0,
\]

where \( d_F := \min_{S^2}(\frac{\lambda_1 + \lambda_2}{2})/\max_{S^2}(\frac{\lambda_2^3}{\lambda_1}) \). Then \( X \) cannot be stable in \( D \).

This proposition leads to the following generalization of the Schwarz criterion [1, Theorem 2.7] for instability of minimal surfaces:

**Theorem 4.5.** Let \( X : M \to \mathbb{R}^3 \) be an \( F \)-minimal surface. If \( N : \overline{D} \to S^2 \) is a branched covering of \( N(D) \) and if the first eigenvalue of \( N(D) \) for \( \Delta_{S^2} \) is smaller than \( 2d_F \), then \( D \) cannot be a stable domain of \( X \).

**Proof.** As in the Proof of Theorem 1.1, we consider an eigenfunction \( u > 0 \) of \( N(D) \) for \( \Delta_{S^2} \), i.e.,

\[
\Delta_{S^2} u + \mu_1 u = 0 \text{ in } N(D) \\
\mu_1 u = 0 \text{ on } \partial N(D).
\]

Lifting \( u \) on \( D \) we obtain the equation

\[
\int_D |\nabla_N v|^2_N - 2d_F v^2 \, dA_N = (\mu_1 - 2d_F) \int_D v^2 \, dA_N
\]

for the function \( v := u \circ N \). By assumption we have \( \mu_1 < 2d_F \) and Proposition 4.4 completes the proof. □

Let us conclude this section applying the main result Theorem 1.1 to a certain class of integrands \( F^\beta \), where

\[
F^\beta(z) = \sqrt{\beta}|z_1|^2 + |z_2|^2 + |z_3|^2,
\]

and \( \beta > 0 \).

The positive eigenvalues of \( F^\beta_{zz}(z) \) for \( z \in S^2 \) are given by \( 1/F^\beta(z) \) and \( \beta/[F^\beta(z)]^3 \). For \( \beta \geq 1 \) one has \( \lambda_1 = 1/F^\beta(z) \) and \( \lambda_2 = \beta/[F^\beta(z)]^3 \). In case \( \beta < 1 \) we get \( \lambda_1 = \beta/[F^\beta(z)]^3 \) and \( \lambda_2 = 1/F^\beta(z) \). This leads to

\[
\max_{S^2} \left( \frac{\lambda_1 + \lambda_2}{2} \right) = \begin{cases} (1 + \beta)/2 & : \beta \geq 1 \\ 1/\sqrt{\beta} & : \beta < 1 \end{cases}
\]

\(^1\)This example was added in proof.
The constant $c_{F\beta}$ (see Proposition 4.2) is dependent on $\beta$ in the following way:

$$c_{F\beta} = \begin{cases} \beta (1 + \beta/2) & : \beta \geq 1 \\ 1/\beta^2 & : \beta < 1. \end{cases}$$

Thus we see $\lim_{\beta \to 0} c_{F\beta} = \lim_{\beta \to \infty} c_{F\beta} = \infty$ and therefore the corresponding area $a_{F\beta}$ (see Theorem 1.1) tends to zero in both cases. Thus for extreme anisotropic integrands the condition for stability is very strong. This is also true for the instability criterion Theorem 4.5 because in this case $\lim_{\beta \to 0} d_{F\beta} = \lim_{\beta \to \infty} d_{F\beta} = 0$.

5. A topological property of the normal of $F$-minimal surfaces.

In the Proof of Theorem 1.1 we used the open-mapping property of the normal $N$ of an $F$-minimal surface. In fact, more is true:

**Theorem 5.1.** The Gauß-map of an $F$-minimal surface $X : M \to \mathbb{R}^3$ is a local branched covering for elliptic integrands $F$.

The above theorem follows from:

**Proposition 5.2.** Let $\omega : B_1 \to \mathbb{C}$, $B_1$ the unit disc in $\mathbb{C}$, be a bounded solution of

$$(9) \quad |\nabla \omega|^2 \leq 2c J_\omega, \ c \geq 1,$$

where $J_\omega$ is the Jacobian of $\omega$. The mapping $\omega$ is a local branched covering if $\# \{|\nabla \omega| = 0\} < \infty$.

For the proof of Theorem 5.1 we have to justify, that for all $p \in M$ there is a chart $x : U(p) \to B_1(0)$ such that $\{K = 0\} \cap U$ consists of only one point and that the stereographic projection of $N \circ x^{-1}$ is a solution of (9). This fact is a result of Sauvigny [13]. Now we can apply Proposition 5.2 and for the completion of the proof of Theorem 5.1 it remains to show Proposition 5.2.

**Proof of Proposition 5.2.** First we see that $|\omega_z(z)|^2 - |\omega_z(z)|^2 > 0$ a.e., where $\frac{\partial}{\partial z}, \frac{\partial}{\partial \bar{z}}$ are the Wirtinger-derivatives and $z = x + iy \in B_1$. This inequality is true because of (9). Now, the differential inequality shows that

$$\mu := \frac{\omega_z}{\omega_z}$$

is an $L^\infty$-function with $||\mu||_\infty \leq k < 1$, where $\frac{1+4z^2}{1+4z^2} := c$. By a result of Morrey [8, p. 204], there is a homeomorphism $\lambda : \overline{B_1} \to \overline{B_1}$ of class $H^{1,2}$
satisfying
\[
\lambda(0) = 0 \\
\lambda_\tau = \mu \lambda_z \quad \text{a.e. on} \ B_1.
\]
Now we want to show that \( \varphi := \omega \circ \lambda^{-1} \) is a holomorphic mapping.

To this aim, we chose a point \( z_0 \in \{ |\nabla \omega| \neq 0 \} \) and an open set \( U, z_0 \in U \), such that \( \omega|_U \) is a diffeomorphism. Then one can define
\[
\Phi := \lambda \circ (\omega|_U)^{-1} : \omega(U) \to \lambda(U).
\]
Setting \( V := \omega(U) \) we see that \( \Phi \in H^{1,2}(V, \mathbb{C}) \). With \( \tau := (\omega|_U)^{-1} \) the following equation holds a.e.:
\[
\Phi_\tau = \lambda_\tau \tau_\zeta + \lambda_\tau \tau_\bar{\zeta} = \lambda_\zeta (\tau_\zeta + \mu \tau_\bar{\zeta}), \quad \zeta \in V.
\]
By differentiation of the identity \( \zeta = \omega(\tau(\zeta)) \) we obtain:
\[
1 = \omega_\zeta (\tau_\zeta + \mu \tau_\bar{\zeta}), \\
0 = \omega_\zeta (\tau_\bar{\zeta} + \mu \tau_\zeta).
\]
These equations show \( \tau_\zeta + \mu \tau_\bar{\zeta} = 0 \) and therefore \( \Phi_\tau(\zeta) = 0 \) for almost all \( \zeta \in V \). Thus \( \Phi \) is a holomorphic and injective mapping. For this reason, \( \Phi \) is a diffeomorphism and \( \varphi \) is holomorphic on \( \lambda(\{ |\nabla \omega| \neq 0 \}) \). The proof is complete because the set \( \lambda(\{ |\nabla \omega| = 0 \}) \) consists only of removable singularities.

\textbf{Acknowledgements.} The author would like to thank the Sonderforschungsbereich 256 at the University of Bonn for its generous support and in particular S. Hildebrandt and M. Rumpf who made possible this publication.

\textbf{References}

J. Math., 98 (1976), 515-528, MR 54 #1292, Zbl 0332.53006.


Received August 29, 2001.

GERHARD-MERCATOR-UNIVERSITÄT DUISBURG
FACHBEREICH MATHEMATIK
LOTTHARSTRASSE 63/65
47048 DUISBURG
GERMANY

E-mail address: clarenz@math.uni-duisburg.de
PARABOLIC SUBGROUPS OF ARTIN GROUPS
OF TYPE FC

Eddy Godelle

Let \((A, S)\) be an Artin group of type FC and \(A_T\) a standard parabolic subgroup of \(A\). We use combinatorial tools to show that the normalizer of \(A_T\), the commensurator of \(A_T\), and the product of the quasi-centralizer of \(A_T\) by \(A_T\) are equal. Furthermore, we show that the centralizer and the quasi-centralizer of \(A_T\) in \(A\) are generated by their intersections with the monoid \(A^+\).

0. Introduction.

Let \(S\) be a finite set and \(M = (m_{s,t})_{s,t \in S}\) a symmetric matrix with \(m_{s,s} = 1\) for \(s \in S\) and \(m_{s,t} \in \mathbb{N} - \{0, 1\} \cup \{\infty\}\) for \(s \neq t\) in \(S\). An Artin-Tits system associated to \(M\) is the pair \((A_S, S)\) where \(A_S\) is the group defined by the presentation

\[ A_S = \left\langle S \mid \frac{sts\ldots}{m_{s,t} \text{ terms}} = \frac{tst\ldots}{m_{s,t} \text{ terms}}; \forall s, t \in S, s \neq t \text{ and } m_{s,t} \neq \infty \right\rangle. \]

The group \(A_S\) is called an Artin group and relations \(\frac{sts\ldots}{m_{s,t} \text{ terms}} = \frac{tst\ldots}{m_{s,t} \text{ terms}}\) are called braid relations. For instance, if \(S = \{s_1, \ldots, s_n\}\) with \(m_{s_i,s_j} = 3\) for \(|i - j| = 1\) and \(m_{s_i,s_j} = 2\) otherwise, then the associated Artin group is the braid group. We denote by \(A_S^+\) the submonoid of \(A_S\) generated by \(S\). This monoid \(A_S^+\) has the same presentation as the group \(A_S\), considered as a monoid presentation ([11]). When we add relations \(s^2 = 1\) to the presentation of \(A_S\) we obtain the Coxeter group \(W_S\) associated to \(A_S\). We say that \(A_S\) is spherical if \(W_S\) is finite. The matrix \(M\) may be represented by a graph denoted by \(\Gamma_S\), whose set of vertices is \(S\) and where an edge joins two vertices if \(m_{s,t} \geq 3\); these edges are labelled by \(m_{s,t}\) if \(m_{s,t} \geq 4\). We say that \(A_S\) (or simply \(S\)) is indecomposable if the graph \(\Gamma_S\) is connected. A subgroup \(A_T\) of \(A_S\) generated by a part \(T\) of \(S\) is called a standard parabolic subgroup, and a subgroup of \(A_S\) conjugate to a standard parabolic subgroup is called a parabolic subgroup. Van Der Lek showed ([14]) that \((A_T, T)\) is canonically isomorphic to the Artin-Tits system associated to the matrix \((m_{s,t})_{s,t \in T}\); its graph \(\Gamma_T\) is the full subgraph of \(\Gamma_S\) generated by \(T\). The
indecomposable components of $S$ are the maximal subsets of $S$ which are indecomposable.

One says that an Artin-Tits system $(A_S, S)$ (or simply an Artin group $A_S$) is of type FC if the following assertion is true:

$$\forall T \subset S, (\forall s, t \in T, m_{s,t} \neq \infty \Rightarrow A_T \text{ is spherical}).$$

For instance, the Artin group of the following graph is of type FC;

![Graph](image)

If $T$ is a subset of $S$ we call centralizer (resp. quasi-centralizer, normaliser, commensurator) of $A_T$ in $A_S$ the set

$$Z_{A_S}(A_T) = \{g \in A_S | \forall s \in T, gs = sg\},$$

$$QZ_{A_S}(A_T) = \{g \in A_S | gT = Tg\},$$

$$N_{A_S}(A_T) = \{g \in A_S | gT \subset A_Tg\},$$

$$\text{Com}_{A_S}(A_T) = \{g \in A_S | gA_Tg^{-1} \cap A_T \text{ has finite index in both } A_T \text{ and } gA_Tg^{-1}\}$$

respectively. These sets are subgroups of $A_S$.

The first of the three main theorems we will prove is the following:

**Theorem 0.1.** Let $(A_S, S)$ be an Artin-Tits system of type FC and $X \subset S$; then

$$\text{Com}_{A_S}(A_X) = N_{A_S}(A_X) = A_X \cdot QZ_{A_S}(A_X).$$

This result was first proved by Rolfsen ([12]) in the case of braid groups; Paris ([10]) proved it when $A_S$ is a spherical Artin group and $T$ is indecomposable, finally in [8] we proved the result for any $T$ when $S$ is spherical.

In [6], the quasi-centralizer in the braid group of a part $T$ of $S$ is geometrically described thanks to the notion of ribbon; this notion was first generalised from a combinatorial viewpoint in [10] (and called conjugator) to a general Artin group and indecomposable part $T$, and finally generalised in [8] for any part $T$. The right viewpoint is to use the categorical language and to see the quasi-centralizer (and centralizer) as a set of morphisms in a groupoid.

Recall that $(A_S, S)$ is spherical if and only if $S$ has a lcm in $A_S^+$; in that case, this lcm is denoted by $\Delta_S$.

**Definition 0.2.** Let $(A_S, S)$ be an Artin-Tits system.

(i) We define the groupoid $\text{Conj}(S)$ as follows:

(a) Objects of $\text{Conj}(S)$ are subsets of $S$;

(b) the set $\text{Conj}(S; X, Y)$ of morphisms from $X$ to $Y$ is in bijection with the set

$$\{g \in A_S | gXg^{-1} = Y\};$$
(c) the composition of morphisms is defined by the product in $A_S$:
\[ g \circ f = gf. \]

(ii) Let $X, Y \subset S$; we say that $w \in \text{Conj}(S; X, Y)$ is a positive elementary $Y$-ribbon-$X$ ([10, 8]) if:
(a) $w = \Delta_{X'}$ for $X'$ an indecomposable component of $X$ or,
(b) there exists $t \in S$ such that the indecomposable component $X'$ of $X \cup \{t\}$ containing $t$ is spherical and $w = \Delta_{X'} \Delta_{X'\setminus\{t\}}^{-1}$.

We say that $w \in \text{Conj}(S; X, Y)$ is an elementary $Y$-ribbon-$X$ if it is a positive elementary ribbon or $w^{-1}$ is a positive elementary $X$-ribbon-$Y$.

(iii) We denote Ribb($S$) the smallest subcategory of Conj($S$) which has the same objects and which contains the elementary ribbons; the set of morphisms from $X$ to $Y$ in Ribb($S$) is denoted Ribb($S; X, Y$) and its elements are called $Y$-ribbon-$X$.

Note that in Case (ii)(a), $X = Y$ and that in Case (ii)(b) there exists $u \in S$ such that $X \cup \{u\} = Y \cup \{t\}$.

The second main theorem of this article is the following:

**Theorem 0.3.** Let $(A_S, S)$ be an Artin-Tits system of type FC; then the category Conj($A_S, S$) is generated by the elementary ribbons; that is Conj($S$) = Ribb($S$).

This result was proved by Paris in [10] for spherical Artin groups. In [8] we proved a similar result in all Artin monoids; in that case, generators are the positive elementary ribbons.

**Corollary 0.4.** Let $(A_S, S)$ be an Artin-Tits system of type FC and $X \subset S$; then
\[ w \in QZ_{A_S}(A_X) \iff w = w_n \ldots w_1 \text{ with } w_i \text{ an elementary } X_i\text{-ribbon-}X_{i-1} \text{ where } X_0 = X_n = X \]
and $QZ_{A_S}(A_X)$ is the subgroup of $A_S$ generated by $QZ_{A_S}(A_X) \cap A^+_S$.

**Corollary 0.5.** Let $(A_S, S)$ be an Artin-Tits system of type FC and $X \subset S$; then $Z_{A_S}(A_X)$ is the subgroup of $A_S$ generated by $Z_{A_S}(A_X) \cap A_S^+$.

In Section 1, we recall relevant facts on Artin groups of type FC, on Artin monoids and define some useful notations; in Section 2 we look at the spherical case and in Section 3, we prove the main results.

1. Preliminaries.

In this part we assume that $(A, S)$ is an Artin-Tits system associated to the matrix $M = (m_{s,t})_{s,t \in S}$.
Lemma 1.1.

(i) ([2, 9]). $A^+$ is left (resp. right) cancellable and every pair $a, b \in A$ has a left (resp. right) gcd, denoted $a \wedge_l b$ (resp. $a \wedge_r b$).

(ii) ([14, Theorem II.4.13]). Let $T$ be a nonempty subset of $S$. Then the subgroup $(\langle T \rangle_A, T)$ of $A$ is canonically isomorphic to the Artin group $A_T$ associated to the matrix $(m_{s,t})_{s,t \in T}$. Furthermore, if $T'$ is another subset of $S$, then $A_T \cap A_{T'} = A_{T \cap T'}$ with the notation $A_\emptyset = \{1\}$.

Lemma 1.2. Assume that $A$ is spherical and let $a, b \in A^+$; then

(i) ([2]). $a, b$ have a left lcm (resp. right lcm) in $A^+$ denoted $a \lor_l b$ (resp. $a \lor_r b$).

(ii) ([5, Paragraph 4]). Let $g \in A$; $g$ can be written $g = g_1 \Delta^n_S$ with $g_1 \in A^+$, and $n \in \mathbb{Z}$.

(iii) ([3, Theorem 2.6] and [4, Lemma 4.4]). Let $g \in A$; there exists unique $a, b \in A^+$ such that $a \wedge_r b = 1$ and $g = ab^{-1}$. Furthermore, if $c \in A^+$ such that $gc \in A^+$ then $c = bc'$ for some $c' \in A^+$.

We call the decomposition $g = ab^{-1}$ of (iii) the (right) orthogonal splitting of $g$. In a similar way one can define the left orthogonal splitting of $g$.

Lemma 1.3 ([8, Corollary 4.4.6]). Let $(A, S)$ be a spherical Artin group and let $s, t \in S$, $g \in A$ and $j \in \mathbb{N}^*$ be such that $s^j g = gt^j$. Then:

(i) $sg = gt$;

(ii) if $w = ab^{-1}$ is the orthogonal splitting of $g$, then

$$\begin{align*}
\text{for some } u \in S.
\end{align*}$$

\begin{align*}
\{ & sa = au, \\
& tb = bu \}
\end{align*}

Notation 1.4. Let $(A, S)$ be an Artin group and let $X \subset S$.

(i) We denote by $X_s$ the union of the spherical indecomposable components of $X$ and by $X_{as}$ the complement $X - X_s$.

(ii) We denote by $X^\perp$ the set $\{s \in S \mid \forall t \in X, \ m_{s,t} = 2\}$; we have $X \cap X^\perp = \emptyset$.

(iii) If $Y$ is another subset of $S$ we write $X \cup Y = X \oplus Y$ if $Y \subset X^\perp$. In particular, $X = X_s \oplus X_{as}$.

(iv) If $s \in X$, we denote by $X(s)$ the indecomposable component of $X$ which contains $s$.

In the following we write $X_{as}^\perp$ for $(X_{as})^\perp$.

To prove our main results we need to introduce the following notations for a new family of subcategories of Conj$(S)$ which generalises Ribb$(S)$; we only give notations for their morphisms.
Notation 1.5. Let \((A_S, S)\) be an Artin group, and \(T \subset S\). Consider the smallest subcategory of \(\text{Conj}(S)\) which has the same objects as \(\text{Conj}(S)\) and which contains the elementary ribbons which are in \(AT\). For \(X,Y \subset Y\), we denote by \(\text{Ribb}(T;X,Y)\) the set of morphisms from \(X\) to \(Y\) in this subcategory. They are \(Y\)-ribbons-\(X\).

1.1. Artin-Tits system of type FC. We assume in this section that \((A,S)\) is of type FC. Recall that Artin groups of type FC have been defined in the introduction. Most facts on Artin groups of type FC in this part are proved in [1].

Proposition 1.6 ([13, Theorem 1]). Let \(G = G_1 *_H G_2\) the amalgamated product of groups \(G_1\) and \(G_2\) over \(H\). Let \(C_1, C_2\) be transversals of \(G_1/H\) and \(G_2/H\) respectively which contain 1. For all \(x \in G\), there exists a unique sequence \((x_1, \ldots, x_n, h)\) such that \(x = x_1 \ldots x_n h\) with \(h \in H\), and where the \(x_i\) are in \(C_1 \cup C_2\) with \(x_i \neq x_{i+1}\) not in the same transversal.

We will call \((x_1, \ldots, x_n, h)\) the amalgam normal form of \(x\) relative to the amalgamated product \(G_1 *_H G_2\) and we set \(|x| = n\). We have then \(|x| = 0\) if and only if \(x \in H\).

Corollary 1.7 ([1, Corollary 1]). Let \(G = G_1 *_H G_2\) and \(g,c \in G\). We denote by \((g_1, \ldots, g_n, h)\) the amalgam normal form of \(g\). Assume that \(g_n \in C_1\) and \(|c| \leq 1\), then: The amalgam normal form of \(gc\) is

\[
\begin{cases}
(g_1, \ldots, g_n, g_{n+1}, h') & \text{if } c \in G_2 - H, \\
(g_1, \ldots, g_{n-1}, g'_n, h) & \text{if } c \in G_1 - (g_n h)^{-1} H, \\
(g_1, \ldots, g_{n-1}, h') & \text{if } c \in (g_n h)^{-1} H,
\end{cases}
\]

where \((g_{n+1}, h')\) is the amalgam normal form of \(hc\) in the first case, \((g'_n, h')\) is the amalgam normal form of \(g_n hc\) in the second case and \(h' = g_n hc\) in the third case.

Corollary 1.8. Let \(w = v_1 \ldots v_m \in G\) such that \(v_{2j} \in G_2 - H\) and \(v_{2j+1} \in G_1 - H\) for \(j \in \{0, \ldots, \lfloor \frac{m}{2} \rfloor\}\). If we denote by \((w_1, \ldots, w_n, h)\) the amalgam normal form of \(w\), then one has

\[
\begin{align*}
m &= n, \\
v_1 &= w_1 h_1 \text{ with } h_1 \in H, \\
h_{i-1}v_i &= w_i h_i \text{ with } i \in \{2, \ldots, n\} \text{ with } h_i \in H, \\
h_n &= h.
\end{align*}
\]

Proposition 1.9 ([1, Proposition 2]).

(i) Let \(s_1, s_2\) be in \(S\) be such that \(m_{s_1, s_2} = \infty\). Let \(A_1 = A_{S-\{s_1\}}, A_2 = A_{S-\{s_2\}}\) and \(A_{1,2} = A_{S-\{s_1, s_2\}}\), then the group \(A\) is the amalgamated product of \(A_1\) and \(A_2\) over \(A_{1,2}\), that is \(A = A_{1,2} * A_{1,2} A_2\).

(ii) The set of Artin groups of type FC is the smallest class of Artin groups which is closed under amalgamation over standard parabolic subgroups and which contains spherical Artin groups.
Proposition 1.10 ([1, Theorem 2]). Let $T \subset S$. There exists a function $m_T : A \to A$ such that for all $w \in A$ one has:

(i) $m_T(w) \in wA_T$;
(ii) for all $v \in wA_T$, $m_T(v) = m_T(w)$;
(iii) if $w \in A_U$ for $U \subset S$, then $m_T(w) \in A_U$.

The function $m_T$ gives a special representative of each coset of $A/A_T$.

Notation 1.11. Assume that $(A,S)$ is not spherical and fix $s_1, s_2 \in S$ such that $m_{s_1,s_2} = \infty$. We set $A_1 = A_{s_1}$, $A_2 = A_{s_2}$ with $S_1 = S - \{s_1\}$, $S_2 = S - \{s_2\}$ and $S_{1,2} = S - \{s_1, s_2\}$. Then we have $A = A_1 \ast A_{1,2} A_2$. Transversals of $A_1/A_{1,2}$ and $A_2/A_{1,2}$ are transversals $C_1, C_2$ respectively induced by $m_{s_1, s_2}$.

Corollary 1.12. Assume that $(A,S)$ is not spherical and let $s_1, s_2 \in S$ with $m_{s_1, s_2} = \infty$; one has $A = A_1 \ast A_{1,2} A_2$ with Notation 1.11. If $w \in A_T$ for $T \subset S$ then the amalgam normal form of $w$ has its terms in $A_T$.

1.2. Artin monoids.

Definition 1.13. Let $(A_S, S)$ be an Artin-Tits system.

(i) We define the small category $\text{Conj}^+(S)$ as follows:
   (a) Objects of $\text{Conj}^+(S)$ are subsets of $S$;
   (b) the set $\text{Conj}^+(S; X, Y)$ of morphisms from $X$ to $Y$ is in bijection with the set
       \[ \{ g \in A^+ \mid gXg^{-1} = Y \} \];
   (c) the composition of morphisms is defined by the product in $A^+$:
       \[ g \circ f = gf. \]

(ii) We denote $\text{Ribb}^+(S)$ the smallest subcategory of $\text{Conj}^+(S)$ which has the same objects and which contains the positive elementary ribbons (see 0.2); the set of morphisms from $X$ to $Y$ in $\text{Ribb}^+(S)$ is denoted $\text{Ribb}^+(S; X, Y)$ and its elements are called positive $Y$-ribbon-$X$.

Categories $\text{Conj}^+(S)$ and $\text{Ribb}^+(S)$ are clearly subcategories of $\text{Conj}(S)$ and $\text{Ribb}(S)$ respectively.

In the following, we will need the following theorem in the spherical case. It is Theorem 0.3 but in the setting of the Artin monoid.

Theorem 1.14 ([10]). Let $(A_S, S)$ be an Artin-Tits system of spherical type; then

\[ \text{Conj}^+(S) = \text{Ribb}^+(S). \]

In fact this theorem is true in any Artin monoid ([8]).
2. The spherical case.

As we said in the introduction, Theorems 0.1 and 0.3 are known in the spherical case. Nevertheless we need to state precise results in the spherical case to prove our theorems in type FC.

**Theorem 2.1.** Let \((A,S)\) be a spherical Artin group and \(X,Y \subset S\). Let \(k \in \mathbb{Z} - \{0\}\) and \(g \in A\). The following are equivalent:

1. \(gA_Xg^{-1} \subset A_Y\);
2. \(g\Delta_X^kg^{-1} \in A_Y\);
3. \(g = yx \text{ with } y \in A_Y\), \(x \in \text{Ribb}(S;X,R)\) for \(R \subset Y\).

**Proof.** It is clear that (3) \(\Rightarrow\) (1) \(\Rightarrow\) (2). For (2) \(\Rightarrow\) (3), the proof is similar to Proposition 3.1 of [7]; thanks to Lemma 1.2(ii), we may assume that \(g \in A^+\) and is \(Y\)-reduced (i.e., not divisible by any \(s \in Y\)); then for all \(s \in X\), we have \(gsg^{-1} = t\) for some \(t \in Y\). Thus \(g \in \text{Ribb}(S;X,R)\) with \(R \subset Y\) by Theorem 1.14. \(\square\)

**Lemma 2.2.** Let \((A,S)\) be a spherical Artin group and let \(X,Y,T \subset S\). Let \(g \in A_T\) be such that \(gA_Xg^{-1} \subset A_Y\). Let \(s \in X - T\); then there exists \(x \in A_{X(s) + \cap T}^+\) and \(y \in A_{Y(T)}^+\) such that \(g = yx\). Furthermore \(X(s) \subset Y\).

**Proof.** Let \(g = a_0b_0^{-1}\) the orthogonal splitting of \(g \in A_T^+\). One has \(a_0 = a_1a\) where \(a\) is \(Y\)-reduced and \(a_1 \in A_T^{+\cap T}\). In the same way, one has \(b_0 = b_1b\) where \(b\) is \(\{s\}^\perp\)-reduced and \(b_1 \in A_T^{+\cap T}\). We obtain \(ab^{-1}sba^{-1} \in A_Y\); hence \(ab^{-1}sba^{-1} = u^{-1}v\) with \(u \perp v\) in \(A_T^+\). Thus \(b^{-1}s = (ua)^{-1}(va)\) with \(b \perp s\) since \(b \notin A_{S-\{s\}}^+\) and is \(\{s\}^\perp\)-reduced. Thus, there exists \(a\) in \(A^+\) such that \(va = \alpha sb\) and \(ua = \beta b\). This implies that \((b \lor_a a)^{-1}\) divides \(v\) and thus is in \(A_{Y\cap T}\). On the other hand, we have \(ba^{-1} = c^{-1}d\) with \(c = (b \lor_a a)b^{-1} \in A_T^+\) and \(d = ((b \lor_a a)^{-1}) \in A_T^{+\cap Y}\). Thus \(csc^{-1} \in A_Y\). Let \(c = c_2c_1\) with \(c_1 \in A_T^{+\cap T}\) and \(c_2\) reduced-\(\{s\}^\perp\) in \(A_T^+\). Then we have \(c_2sc_2^{-1} \in A_Y\) with \(c_2s \perp c_2\). Thus both \(c_2\) and \(s\) are in \(A_T\). Then \(g = y_0x_0\) with \(x_0 = c_1b_1^{-1} \in A_{T(s)}^{+\cap T}\) and \(y_0 = a_1d^{-1}c_2 \in A_{Y\cap T}\). We have \(x_0A_Xx_0^{-1} \subset A_Y\) with \(x_0 \in A_{T(s)}^{+\cap T}\). If \(x_0 = 1\) or \(X(s) = \{s\}\), the result holds with \(x = x_0\) and \(y = y_0\). Assume \(x_0 \neq 1\) and \(X(s) \neq \{s\}\). Choose \(s' \in X(s) - \{\{s\}^\perp \cup \{s\}\}\) (it exists since \(X(s) \neq \{s\}\)). Applying the argument to \(g' = x_0, T' = \{s\}^\perp \cap T, \) and \(s',\) we obtain \(x_0 = y_1x_1\) with \(x_1 \in A_{T(s')}^{+\cap T}\) and \(y_1 \in A_{Y\cap T}\). Repeating this process yields \(g = y_0 \ldots y_nx_n\) with \(x_n \in A_{T(s)^n_{\cap T}}\) and \(y_0 \ldots y_n \in A_{T\cap Y}\). The process will terminate when either \(x_n = 1\) or

\[
X(s) - \left( \bigcap_{i=0}^{n} \{s(i)\}^\perp \cup \bigcup_{i=0}^{n} \{s(i)\} \right) = \emptyset
\]

which means that \(X(s) = \bigcup_{i=0}^{n} \{s(i)\}^\perp = A_{X(s)^n \cap T}\). In either case, the result follows with \(x = x_n\) and \(y = y_0 \ldots y_n\). \(\square\)
Proposition 2.3. Under the hypotheses of Theorem 2.1, if \( g \in A_T \) for \( T \subset S \), then (1), (2) and (3) are equivalent to

\[
(3') \quad g = yx \text{ with } y \in A_{Y \cap T}, \ x \in \text{Ribb} (T \cap (\bigcup_{s \in X - T} X(s))^{\perp}; X, R) \text{ where } R \text{ is of the form } \bigcup_{s \in X - T} X(s) \oplus T_1 \subset Y \text{ with } T_1 \subset T \cap Y.
\]

Proof. It is clear that \((3') \Rightarrow (3)\). Let us show that \((1) \Rightarrow (3')\) by induction on the cardinal of \( X - T \). It is enough to find \( x, y \) such that \( y \in A_T \) and \( x \in \text{Ribb} (T \cap (\bigcup_{s \in X - T} X(s))^{\perp}; X, R) \) since that implies \( y \in A_T \) and the type of \( R \). If \( X - T = \emptyset \), that is \( X \subset T \), then \( gA_X g^{-1} \subset A_{Y \cap T} \) in \( A_T \) and we apply Theorem 2.1 in \( A_T \). Otherwise Lemma 2.2 proves that for all \( \#(X - T) \geq 1 \) and \( s \in X - T \) then \( g = y_1 x_1 \) with \( x_1 \in A_{X(s)^{\perp} \cap T} \) and \( y_1 \in A_{Y \cap T} \). Thus, in \( A_{X(s)^{\perp}} \) we have \( x_1 (X - X(s)) x_1^{-1} \subset A_{Y \cap X(s)^{\perp}} \) with \( x_1 \in A_{T \cap X(s)^{\perp}} \). We apply the induction hypothesis in \( A_{X(s)^{\perp}} \), after replacing \( g \) by \( x_1, Y \) by \( X - X(s), T \) by \( T \cap X(s)^{\perp} \) and \( Y \) by \( Y \cap X(s)^{\perp} \); we have \( \#((X \cap X(s)^{\perp}) - (T \cap X(s)^{\perp})) < \#(X - T) \) since \( s \notin X \cap X(s)^{\perp} \) and \( s \in X - T \). We get \( x_1 = y_2 x \) with \( x \) in \( \text{Ribb} (T \cap X(s)^{\perp} \cap \bigcup_{u \in X - X(s)^{\perp} - T} X(u)^{\perp}; X - X(s), R_1) \) with \( R_1 \subset Y \) and \( y_2 \in A_Y \). But \( X(s)^{\perp} \cap \bigcup_{u \in X - X(s)^{\perp} - T} X(u)^{\perp} = \bigcup_{u \in X - T} X(u)^{\perp} \). Thus \( g = yx \) with \( y = y_1 y_2 \in A_Y \) and \( x \in \text{Ribb} (T \cap (\bigcup_{u \in X - T} X(u))^{\perp}; X, R) \). \( \square \)

3. Proof of the main results.

Proposition 3.1. Let \((A, S)\) be an Artin group of type FC. Let \( X, Y, T \subset S \) with \( X \) spherical. Let \( k \in \mathbb{Z} - \{0\} \) and \( g \in A_T \). The following are equivalent:

1. \( gA_X g^{-1} \subset A_Y \);
2. \( gA_X^k g^{-1} \in A_Y \);
3. \( g = yx \) with \( y \in A_{Y \cap T} \) and \( x \in \text{Ribb} (T \cap (\bigcup_{s \in X - T} X(s))^{\perp}; X, R) \) for some \( R \subset Y \).

Proof. Implications \((3) \Rightarrow (1) \Rightarrow (2)\) are clear. Let us show that \((2) \Rightarrow (3)\) by induction on the number \( m \) of amalgamations; that is the number of edges in \( \Gamma_S \) labelled with \( \infty \). If \( m = 0 \) then \( A \) is spherical and the result is true by Theorem 2.1 and Proposition 2.3. Assume now that \( m \geq 1 \) and that Proposition 3.1 is true for any Artin group of type FC with a number of amalgamation less than or equal to \( m - 1 \). We choose \( s_1, s_2 \in S \) such that \( m_{s_1, s_2} = \infty \) and we use Notation 1.11. Note that since \( A_X \) is spherical, we have \( X \subset S_1 \) or \( X \subset S_2 \). Denote by \((g_1, \ldots, g_n, h)\) the amalgam normal form of \( g \); elements \( g_i \) and \( h \) are in \( A_T \) by Corollary 1.12. For \( m \) fixed, let us do an induction on \( n \). We have \( g_1 \ldots g_n h A_X^k h^{-1} g_n^{-1} \ldots g_1^{-1} \in A_Y \). If \( n = 0 \) then the formula holds in \( A_1 \) or in \( A_2 \) and we conclude by the induction hypothesis on \( n \) applied in \( A_1 \) or in \( A_2 \). Assume now that \( n \geq 1 \). We may assume without loss of generality that \( g_n \in A_1 \).

If \( X \not\subset A_1 \) then by Corollary 1.8, the amalgam normal form of \( gA_X^k g^{-1} \) is of the shape \((g_1, \ldots, g_n, g'_n, \ldots, g'_{2n-1}, h')\) with \( g'_{n+1} = m_{S_1, 2} (h A_X^k h^{-1}) \).
But \( g \Delta_X^k g^{-1} \in A_Y \) thus \( g_1 \ldots g_n \in A_Y \) by Corollary 1.12 and \( h \Delta_X^k h^{-1} \in A_Y \). Thus, to conclude, we apply the induction hypothesis on \( n \) at rank \( n = 0 \) to \( h \).

If \( X \subset A_1 \) and \( g_n h \Delta_X^k (g_n h)^{-1} \notin A_{1,2} \) then the amalgam normal form of \( g \Delta_X^k g^{-1} \) is of shape \((g_1, \ldots, g_{n-1}, g_n, \ldots, g_{2n-1}, h')\) with

\[ g_n' = ms_{1,2}(g_n h \Delta_X^k (g_n h)^{-1}). \]

But \( g \Delta_X^k g^{-1} \in A_Y \); then we get \( g_1 \ldots g_{n-1} \in A_Y \) by Corollary 1.12 and thus \( g_n h \Delta_X^k (g_n h)^{-1} \in A_Y \); thus we may apply the induction hypothesis on \( m \) in \( A_1 \) at \( g_n h \) and conclude.

If \( X \subset S_1 \) and \( g_n h \Delta_X^k (g_n h)^{-1} \in A_{1,2} \) then by the induction hypothesis on \( m \) applied in \( A_1 \), we get \( g_n h = y_1 x_1 \) with

\[ x_1 \in \text{Ribb} \left( T \cap \left( \bigcup_{t \in X-T} X(t) \right) \downarrow ; X, R \right) \]

for \( R \subset S_{1,2} \) and \( y_1 \in A_{1,2} \cap A_T \). We get \( g_1 \ldots g_{n-1} y_1 \Delta_X^k (g_1 \ldots g_{n-1} y_1)^{-1} \in A_Y \) and by the induction hypothesis on \( n \) applied at rank \( n-1 \) to \( g_1 \ldots g_{n-1} y_1 \in A_T \), we obtain \( g_1 \ldots g_{n-1} y_1 = y x_2 \) with

\[ x_2 \in \text{Ribb} \left( T \cap \left( \bigcup_{t \in X-T} X(t) \right) \downarrow ; R, R_1 \right) \]

for some \( R_1 \subset Y \) and \( y \in A_{Y \cap T} \).

Thus \( g = y x_2 x_1 \) and \( x_2 x_1 \in \text{Ribb} \left( T \cap \left( \bigcup_{t \in X-T} X(t) \right) \downarrow ; X, R_1 \right) \) for \( R_1 \subset Y \).

\[ \square \]

**Theorem 3.2.** Let \((A, S)\) be an Artin group of type FC and let \( X, Y, T \subset S \). Let \( g \in A_T \) and \( k \in \mathbb{Z} \setminus \{0\} \). The following are equivalent:

1. \( g A_X^k g^{-1} \subset A_Y \);
2. \( g \Delta_X^k g^{-1} \in A_Y \), \( g = y x \) with \( x \in A_{X,T}^k \), \( y \in A_{Y \cap T} \) and \( X_{as} \subset Y \);
3. \( g = y x \) with \( y \in A_{Y \cap T} \) and \( x \in \text{Ribb} \left( T \cap X_{as}^+ ; X_s, R \right) \) with \( R \oplus X_{as} \subset Y \).

**Proof.** It is clear that (3) \( \Rightarrow \) (2) and that Proposition 3.1 induces (2) \( \Rightarrow \) (1). Let us show that (1) \( \Rightarrow \) (3). We are carrying out an induction on the number \( r(X) \) of edges in \( \Gamma(X) \) which are labelled with \( \infty \). If \( r(X) = 0 \), then \( X \) is spherical and the result is true by Proposition 3.1. Assume now \( X \) is not spherical (that is \( r(X) \geq 1 \)) and fix \( s_1, s_2 \) in \( X \) such that \( m_{s_1, s_2} = \infty \). We assume that if \((A', S')\) is an Artin group of type FC and \( X', Y', T' \) are three parts of \( S' \) such that \( r(X') < r(X) \) then for all \( g' \) of \( A'_T \), we have:

\[ g'^{-1} A_X^k g' \subset A'_{X,T} \Rightarrow g' = y' x' \text{ where } x' \in \text{Ribb} \left( T' \cap X_{as}'^+ ; X_s', R'_2 \right) \]

and \( y' \in A'_{Y \cap T} \) with \( R'_2 \oplus X_{as}' \subset Y' \).

We have \( A = A_1 \star_{A_{1,2}} A_2 \) with Notation 1.11. Let \((g_1, \ldots, g_n, h)\) be the amalgam normal form of \( g \). The first step is to show that it is enough to prove the result for the case \( n = 0 \). Assume \( n \geq 1 \). Without loss of generality
we may assume that $g_n \in A_1$. Furthermore $g_1 \ldots g_nh_s h^{-1} g_n^{-1} \ldots g_1^{-1} \in A_Y$ since $s_1 \in X$. By Corollary 1.8 we infer that the amalgam normal form of $g_1 \ldots g_nh_s h^{-1} g_n^{-1} \ldots g_1^{-1}$ is of the shape $(g_1, \ldots, g_n, g_{n+1}', \ldots, g_{2n+1}', h')$ and has its terms in $A_Y$. Thus $g_1 \ldots g_n$ is in $A_{Y \cap T}$. We get that $hAxh^{-1}$ is also in $A_Y$. Thus if (1) $\Rightarrow$ (3) for any $g$ such that $n = 0$, the theorem will be proved. Assume $g = h \in A_{12}$. Denote by $T_1$ (resp. $T_2$) the indecomposable component of $X - \{s_1\}$ (resp. $X - \{s_2\}$) which contains $s_2$ (resp. $s_1$). In $A_1$ we have $gAx_{\{s_1\}} g^{-1} \subset A_{Y - \{s_1\}}$ thus by the induction hypothesis, we get $g = y_1 x_1$ with $y_1 \in A_{Y \cap T}$ and $x_1 \in \text{Ribb} \left( T \cap \left( X - \{s_1\}\right)_{\alpha s}; (X - \{s_1\})_s, R_1 \right)$ with $R_1$ in $Y - \{s_1\}$. Furthermore, since $s_2 \notin A_{12}$ and $g \in A_{12}$, we get, either by the induction hypothesis (if $T_1$ is not spherical) or by Proposition 3.1, that we can find $x_1 \in \text{Ribb} \left( T_1^+ \cap T \cap \left( X - \{s_1\}\right)_{\alpha s}; (X - \{s_1\})_s, R_1 \right)$. From $y_1 \in A_Y$, we infer that $x_1 A_1 x_1^{-1}$ is in $A_Y$. We can use the same argument if we replace $g$ by $x_1$ and exchange the roles of $A_1$ and $A_2$; we find, thanks to the inclusion $x_1 A_X x_1^{-1} \subset A_{Y - \{s_2\}}$ in $A_2$, that $x_1 = y_2 x_2$ with $y_2 \in A_{Y \cap T}$ and $x \in \text{Ribb} \left( T_2^+ \cap T \cap \left( X - \{s_1\}\right)_{\alpha s}; (X - \{s_2\})_s, R_2 \right)$ with $R_2$ in $Y$. Finally, since $T_1 \cup T_2 = X(s_1) = X(s_2)$, we get $T_1^+ \cap T \cap \left( X - \{s_1\}\right)_{\alpha s} \supset \left( X - \{s_2\}\right)_s$ and $T \cap X_{\alpha s}$. We get that $g = y_1 y_2 x_2$ with $y_1 y_2 \in A_{Y \cap T}$, $x$ in $\text{Ribb} \left( T \cap X_{\alpha s}; X, R \right)$ since $X_s \subset (X - \{s_2\})_s$ and $R \subset R_2 \subset Y$.

The two following lemmas are used to prove (ii) and the first equality of (i) in Theorem 3.5.

**Lemma 3.3.** Let $(A, S)$ be an Artin group of type FC; then $QZ_{A}(A) = QZ_{A_{S}}(A_{S})$.

**Proof.** It is clear that $QZ_{A}(A) = QZ_{A_{S}}(A_{S}) \cdot QZ_{A_{S}}(A_{S_{\alpha}})$. Then it is enough to prove that $QZ_{A_{S}}(A_{S_{\alpha}}) = \{1\}$ and since it is the product of quasi-centralizers of its indecomposable components, it is enough to show that if $X$ is indecomposable not spherical, then $QZ_{A}(A) = \{1\}$. Let $A$ be such a group and let $g \in QZ_{A}(A)$; choose $T \subset S$ maximal spherical and let $T' = gTg^{-1} \subset S$. Then by Proposition 3.1 applied to the equality $g^{-1}Ag = A_{T'}$, we get that $T = T'$ and $g \in A_T$; thus $g \in QZ_{A_{T}}(A_{T})$. Let $(T_i)_{1 \leq i \leq k}$ be the indecomposable components of $T$ then $QZ_{A_{T}}(A_{T}) = \{\Delta_{T_1}^j \ldots \Delta_{T_k}^j; \forall i, j_i \in \mathbb{Z}\}$. Thus $g = \Delta_{T_1}^j \ldots \Delta_{T_k}^j$ with $j_i \in \mathbb{Z}$ for all $i \in \{1, \ldots, k\}$. Assume that there exists $i \in \{1, \ldots, k\}$ such that $j_i \neq 0$. Since $T$ is maximal spherical and $S$ is indecomposable and not spherical, there exists $s \in T_i$ and $t \in S - T$ such that $m_{s,t} = \infty$. We get $A = A_{S - \{s\}} A_{S - (s, t)} A_{S - \{t\}}$ and $gsg^{-1} = s_1$ with $s_1 \in T_i$; since $j_i \neq 0$, this is impossible by Corollary 1.8. Thus for all $i$, we get $j_i = 0$ and $g = 1$. \hfill $\square$

**Lemma 3.4.** Let $A = A_{1} \ast A_{12}$ be a non-spherical Artin group of type FC in the Notation of 1.11. Let $X \subset S$ be such that $\{s_1, s_2\} \subset X$, let $X_i = X \cap S_i$ for $i \in \{1, 2\}$; then $Com_A(A_X) \cap A_i \subset Com_{A_i}(A_{X_i})$ for $i \in \{1, 2\}$.

**Proof.** By symmetry, it is enough to prove the result for $i = 1$. Let $g \in Com_A(A_X) \cap A_1$. We have to show that $A_{X_1} \cap (gA_{X_1}g^{-1})$ has finite index
in \( A_X \) and in \( gA_Xg^{-1} \). For this, we show that if \( x, y \in A_X \) (resp. \( x, y \in gA_Xg^{-1} \)) have the same image in \( A_X/(A_X \cap gA_Xg^{-1}) \) (resp. \( gA_Xg^{-1}/(A_X \cap gA_Xg^{-1}) \)) then they have the same image in \( A_X/(A_X \cap gA_Xg^{-1}) \) (resp. \( gA_Xg^{-1}/(A_X \cap gA_Xg^{-1}) \)).

Let \( x, y \in A_X \) be such that \( x\alpha = y \) for some \( \alpha \in A_X \). Assume 

\[
\alpha = xy^{-1} \in A_X \cap A_X \cap (gA_Xg^{-1}) = A_X \cap (gA_Xg^{-1}) = A_X \cap (gA_Xg^{-1}),
\]

The last equality come from the fact that \( g \) is in \( A_S \) and that \( A_X \cdot A_S = A_S \cap A_X \). Let \( x, y \in gA_Xg^{-1} \) be such that \( x\alpha = y \) for some \( \alpha \in A_X \cap (A_Xg^{-1}) \). We have by the same arguments that \( \alpha \in A_X \cap (gA_Xg^{-1}) \).

Thus \( g \in \text{Com}_A(A_X) \). 

**Theorem 3.5.** Let \((A, S)\) be an Artin group of type FC. Let \( X \subset S \); then:

1. \( \text{Com}_A(A_X) = N_A(A_X) = A_X \cdot QZ_A(A_X) \);
2. \( \text{Conj}(S; X, Y) = \text{Ribb}(S; X, Y) \subset \text{Ribb}(X_{as}^-, X_s, Y_s) \). Furthermore, if \( \text{Conj}(S; X, Y) \neq \emptyset \) then the inclusion is an equality and \( X_{as} = Y_{as} \).

**Proof.** Thanks to Theorem 3.2 we get the inclusions \( A_X \cdot QZ_A(A_X) \subset N_A(A_X) \subset A_X \cdot \text{Ribb}(X_{as}^-, X_s, X_s) \subset A_X \cdot QZ_A(A_X) \). That proves the second equality of (i): \( N_A(A_X) = A_X \cdot QZ_A(A_X) \). If \( \text{Conj}(S; X, Y) = \emptyset \) then (ii) is clear since \( \text{Ribb}(S; X, Y) \subset \text{Conj}(S; X, Y) \). Assume now that \( \text{Conj}(S; X, Y) \neq \emptyset \). Since \( \text{Conj}(S; X, Y) \subset A_X \cdot \text{Ribb}(X_{as}^-, X_s, X_s) \), one has \( X_{as} = Y_{as} \) and \( \text{Ribb}(X_{as}^-, X_s, Y_s) \subset \text{Conj}(S; X, Y) \); then we get \( \text{Conj}(S; X, Y) = QZ_A(A_X) \cdot \text{Ribb}(X_{as}^-, X_s, Y_s) \). Now, by Lemma 3.3 \( QZ_A(A_X) = QZ_A(A_X) \cdot \text{Ribb}(X_{as}^-, X_s, Y_s) \). Then, we have clearly \( \text{Com}_A(A_X) \supset N_A(A_X) \); if \( X \) is spherical, we prove the other inclusion as in [10] thanks to the implication (2) \( \Rightarrow \) (1) of Proposition 3.1. Let us show that \( \text{Com}_A(A_X) \subset N_A(A_X) \) for \( X \) not spherical. In order to do this, we proceed by induction on the number \( r \) of edges labelled with \( \infty \) in the graph \( \Gamma_X \). Note that for \( r = 0 \), we have that \( X \) is spherical and the result is true in that case. Assume \( r \geq 1 \) and write \( A = A_1 * A_{1,2} A_2 \) such that \( \{ s_1, s_2 \} \in X \) following Notation 1.11. By the induction hypothesis, we have \( \text{Com}_A(A_X \cap S_i) = N_A(A_X \cap S_i) \) for \( i \in \{ 1, 2 \} \).

Let \( g \in \text{Com}_A(A_X) \) and \( (g_1, \ldots, g_n, h) \) its amalgam normal form. There exists \( p \in \mathbb{N} - \{ 0 \} \) such that \( g_s^p g^{-1} \) and \( g_s^p g^{-1} \) are in \( A_X \). If \( n \neq 0 \), we infer from Corollary 1.8 that \( g_1 \ldots g_n \in A_X \) and hence \( h \in \text{Com}_A(A_X) \). Now, for \( g = h \in \text{Com}_A(A_X) \cap A_{1,2} \), we can apply Lemma 3.4; we get that for \( i \in \{ 1, 2 \} \), we have \( h \in \text{Com}_A(A_{X \cap S_i}) \). But by the induction hypothesis, \( \text{Com}_A(A_X \cap S_i) = N_A(A_X \cap S_i) \). Then \( h \in N_A(A_X \cap S_i) \cap N_A(A_X \cap S_2) \subset N_A(A_X) \).

**Corollary 3.6.** Let \((A, S)\) be an Artin groups of type FC. Let \( X \subset S \); then:

1. \( QZ_A(A_X) = \text{Ribb}(X_{as}^-, X_s, X_s) \);
2. \( Z_A(A_X) = (Z_A(A_X) \cap A^+)_A \);
3. \( QZ_A(A_X) = (QZ_A(A_X) \cap A^+)_A \).
Proof. The (i) is a particular case of 3.5(ii); (ii) and (iii) are equivalent by [8] Theorem 4.1.2; furthermore, it is clear that 3.5(ii) implies (iii) thanks to Lemma 1.3. □

Aknowledgements. This work was begun while I was visiting the Laboratory of Topology UMR 5584 CNRS, University of Bourgogne; I am grateful for the attention I received during my stay. Special thanks are due to John Crisp and Luis Paris for their hospitality and useful discussions and remarks. My travel was supported by the GDR Tresses 2105 CNRS.

I thank the referee for his editing work.

References

School of Mathematics & Statistics
University of Sydney
NSW 2006
Australia
E-mail address: eddyg@maths.usyd.edu.au
HASSE PRINCIPLES FOR THE BRAUER GROUPS OF ALGEBRAIC FUNCTION FIELDS OF GENUS ZERO OVER GLOBAL FIELDS

ILSEOP HAN

Let $F$ be a global field with char$(F) \neq 2$ and $K$ an algebraic function field in one variable of genus zero over $F$. In this paper, we investigate two kinds of Hasse principles for Brauer classes on $K$. If $\text{Br}(K)$ is the Brauer group of $K$ and $\text{Br}(K)'$ is the subgroup of $\text{Br}(K)$ whose elements have order relatively prime to char$(F)$, then we precisely determine the kernels of the maps

$$h_1 : \text{Br}(K)' \rightarrow \prod_p \text{Br}((\hat{F}_p)K) \quad \text{and} \quad h_2 : \text{Br}(K) \rightarrow \prod_p \text{Br}((\hat{K}_p)),$$

where $p$ runs over the prime spots of $F$ and $P$ runs over the places of $K$ which are trivial over $F$, and $\hat{F}_p$, $\hat{K}_p$ are the completions at $p$, $P$ respectively. To facilitate the determination of these kernels, we compute the kernel of the map $h : \text{Br}(K) \rightarrow \prod_P \text{Br}(K_{V_P})$ where $V_P$ is the residue field with respect to $P$ and show that the kernels of these three maps coincide. We then consider a more general version of the maps above by describing the 2-torsion subgroup of the kernel of $h_1$ when a finite number of prime spots in the product are omitted.

1. Introduction.

Let $F$ be a global field with char$(F) \neq 2$. By a prime spot on $F$, we mean an equivalence class of discrete valuations on $F$ or an equivalence class of archimedean absolute values on $F$. Define

$$P(F) = \{p \mid p \text{ is a prime spot of } F\}.$$

For $p \in P(F)$ let $\hat{F}_p$ denote the corresponding completion of $F$. These fields $\hat{F}_p$ are the local objects, with relatively easy arithmetic properties. Local global principles allow us to understand properties of $F$ in terms of those over all the $\hat{F}_p$.

Let $K$ be an algebraic function field in one variable over a field $F$. By a place of $K/F$, we mean a normalized discrete valuation on $K$ which is trivial.
on $F^* = F - \{0\}$. Define

$$\mathcal{P}(K/F) = \{P \mid P \text{ is a place of } K/F\}.$$ 

We denote by $\widehat{K}_P$ the completion of $K$ with respect to $P \in \mathcal{P}(K/F)$. Another type of local global principle is to get information about $K$ via the family of $\widehat{K}_P$ for $P \in \mathcal{P}(K/F)$. We will be particularly interested in the case where $K$ has genus 0. Then $K$ has the form $K = F(x, \sqrt{ax^2 + b})$ where $a, b \in F^*$ and $x$ is transcendental over $F$. Since this $K$ is determined up to isomorphism by the quaternion algebra $Q = (a, b/F)$, we will write $F(Q)$ for $K$.

If $\text{Br}(k)$ denotes the Brauer group of a field $k$, then the classical Hasse principle for the Brauer group of a global field $F$ states that the map

$$\text{Br}(F) \rightarrow \prod_{P \in P(F)} \text{Br}(\widehat{F}_P)$$ 

is injective. This is a local global principle, since it says that a central simple algebra $A$ over $F$ is determined by its extensions $A \otimes_F \widehat{F}_P$ over $\widehat{F}_P$ as $P$ ranges over $P(F)$.

One could ask whether there is a corresponding Hasse principle using the $K \otimes_F \widehat{F}_p$, $p \in P(F)$, for the Brauer group of an algebraic function field $K$ over a global field $F$. However, such a Hasse principle no longer holds. The first counterexample was given by Witt (see [Wi, p. 466]) in 1934, taken from an algebraic function field $K$ of genus 0 over $\mathbb{Q}$. He showed that if $Q$ and $Q'$ are quaternion algebras over $\mathbb{Q}$, which are nonsplit just at $p_1, p_2, p_3, p_4 \in P(\mathbb{Q})$ and $p_1, p_2$ respectively, then $Q' \otimes_{\mathbb{Q}} \mathbb{Q}_p(Q)$ is split for every $p \in P(\mathbb{Q})$ although $Q' \otimes_{\mathbb{Q}} \mathbb{Q}(Q)$ is nonsplit. Further, he pointed out that $Q' \otimes_{\mathbb{Q}} \mathbb{Q}_P$ is split for every $P \in P(\mathbb{Q}(Q)/\mathbb{Q})$.

Let $\mathcal{C}$ be an irreducible nonsingular projective curve over a global field $F$. Considering the function field $K = F(\mathcal{C})$ of the curve $\mathcal{C}$ over $F$, one can ask about two kinds of possible Hasse principles for the Brauer group of $K$. One corresponds to the map $h_1$ and the other corresponds to the map $h_2$ below:

$$h_1 : \text{Br}(K)' \rightarrow \prod_{p \in P(F)} \text{Br}(\widehat{F}_p(\mathcal{C}))$$

$$h_2 : \text{Br}(K) \rightarrow \prod_{P \in \mathcal{P}(K/F)} \text{Br}(\widehat{K}_P).$$

Here, $\text{Br}(K)'$ denotes the subgroup of $\text{Br}(K)$ consisting of those $[B]$ with the exponent of $B$ relatively prime to $\text{char}(F)$. So, if $\text{char}(F) = 0$, then $\text{Br}(K)' = \text{Br}(K)$.

In this paper, we consider the case of $K$ having genus 0, putting $K = F(Q)$, and give explicit description of the kernels of the maps $h_1$ and $h_2$ above, which in fact coincide as shown in Theorem 4.4 and Theorem 4.5.
Nontrivial kernels of these maps could be called the obstruction to a Hasse principle.

In order to facilitate the determination of these kernels, we compute, in Theorem 3.7, the kernel of the map

$$h : \text{Br}(K) \longrightarrow \prod_{P \in \mathcal{P}(K/F)} \text{Br}(K\overline{v}_P)$$

where $\overline{v}_P$ is the residue field with respect to $P \in \mathcal{P}(K/F)$. The kernel of $h$ turns out to coincide with those of $h_1$ and $h_2$, and it has a direct description in terms of the quaternion algebra $Q$.

Finally, we give a relative version of the facts shown above. In Theorem 5.8, we completely describe the 2-torsion subgroup of the kernel of $h_1$ when a finite number of prime spots are omitted from $P(F)$. In the case that the prime spots deleted from $P(F)$ contain those $p \in P(F)$ such that $Q \otimes_F \hat{F}_p$ is nonsplit, this part of $\ker(h_1)$ can be expressed in terms of quaternion algebras over $K$ together with the images in $\text{Br}(K)$ of cyclic algebras of exponent 4 over $F$. Otherwise, we can describe the kernel as the intersection of the relative Brauer groups of some quadratic residue fields over $K$ so that it can be all expressed in terms of quaternion algebras over $K$.

Parimala and Sujatha considered in [PS] the kernels of similar maps for the function field of a curve with a rational point. These results do not apply for nonrational function fields of genus 0 since these are function fields of anisotropic conics, which have no rational points. On the other hand, they pointed out that the kernel of $h_1$ is nontrivial for the function field of genus 0 associated with a quaternion algebra which is locally split at 4 prime spots or more.

Kato considered in [Ka, Theorem 0.8(2)] a Hasse principle for $H^3$ cohomology groups analogous to $h_1$ for $H^2$ cohomology groups; he showed that the corresponding map from $H^3(K, \mathbb{Z}/2\mathbb{Z})$ is injective. However, we will see that the map $h_1$ is not injective in general, even when restricted to the 2-torsion of $\text{Br}(K)$, $2\text{Br}(K) = H^2(K, \mathbb{Z}/2\mathbb{Z})$.

This work was motivated by and arose in connection with the author’s work on tractability of algebraic function fields in one variable of genus 0 over global fields (see Remark 5.10). Especially, Chapter 3 is based on a part of the author’s Ph.D. thesis work. We would like to thank Prof. A. Wadsworth at UCSD for providing invaluable guidance and the referee for making helpful comments, which improved the final version of this paper.

## 2. Preliminaries.

In this section, we will briefly review basic facts on the algebraic function fields in one variable of genus 0, which are associated with quaternion algebras.
Let \( F \) be a field with \( \text{char}(F) \neq 2 \). (Throughout, all fields are assumed to have characteristic not equal to 2.) Let \( K \) be an algebraic function field in one variable of genus 0 over \( F \). It is known (cf. [Ar, Theorem 6, p. 302]) that such a \( K \) has a genus 0 if and only if \( K \) is of the form \( F(x, \sqrt{ax^2 + b}) \), where \( a, b \in F^* = F - \{0\} \) and \( x \) is transcendental over \( F \). Let \( (a,b/F) \) denote the 4-dimensional quaternion algebra over \( F \) with \( F \)-base 1, \( i, j, k \), such that \( i^2 = a, \ j^2 = b, \) and \( ij = -ji = k \). Then \( K = F(x, \sqrt{ax^2 + b}) \) is determined by a quaternion algebra \( Q = (a,b/F) \) since \( K \) is isomorphic to the function field of a conic determined by norm form on the pure part of \( Q \). Thus \( K \) will be also denoted by \( F(Q) \). Recall (cf. [Wi, Satz, p. 464 and Satz, p. 465]) that for two quaternion algebras \( Q \) and \( Q' \), \( Q \sim Q' \) as algebras if and only if \( F(Q) \sim F(Q') \) as fields.

For each place \( P \in \mathbb{P}(K/F) \), let \( V_P \) be the associated discrete valuation ring. Recall (e.g., [DI, Lemma 2.2, p. 136]) that the restriction map \( \text{Br}(V_P) \to \text{Br}(K) \) induced by the inclusion \( V_P \hookrightarrow K \) is injective. If \( V_P \) denotes the residue field of \( V_P \), then \( V_P \) is a finite degree extension of \( F \). The degree of \( P \) is defined by \( \text{deg}(P) = [V_P:F] \). Now, let \( F_{\text{sep}} \) be the separable closure of \( F \). We will denote by \( G_P = \text{Gal}(F_{\text{sep}}/V_P) \) the absolute Galois group of \( V_P \) and by \( X(G_P) = \text{Hom}_c(G_P, \mathbb{Q}/\mathbb{Z}) \) the (continuous) character group of \( G_P \). The following result, due to Scharlau, will be essential to our study of algebraic function fields of genus 0.

**Proposition 2.1** (Scharlau). Let \( F \) be any field. For a quaternion algebra \( Q = (a,b/F) \), let \( K = F(Q) = F(x, \sqrt{ax^2 + b}) \). Then the following sequence is exact:

\[
0 \longrightarrow \{ [F], [Q] \} \longrightarrow \text{Br}(F) \overset{\alpha}{\longrightarrow} \text{Br}(F_{\text{sep}} \cdot K/K) \overset{\beta}{\longrightarrow} \bigoplus_{P \in \mathbb{P}(K/F)} X(G_P).
\]

For details of Proposition 2.1, see [Sc, p. 5].

The following well-known lemma will be useful to determine when an algebraic function field of genus 0 is a rational function field.

**Lemma 2.2.** Let \( F \) be any field. For a quaternion algebra \( Q \) over \( F \), let \( K = F(Q) \). Then the following conditions are equivalent:

(i) \( Q \) is split over \( F \).

(ii) There exists \( P \in \mathbb{P}(K/F) \) with \( \text{deg}(P) = 1 \).

(iii) \( K \) is purely transcendental over \( F \).

A proof can be found, e.g., in [Wa, p. 747] or in [Ha, Lemma 3.2].

Note that if the quaternion algebra \( Q \) is split over \( F \), then Proposition 2.1 reduces to the Auslander-Brumer-Faddeev Theorem (cf. [AB], [Fa] or [FS]) in view of Lemma 2.2. This lemma provides a corollary (see [Ha, Corollary 3.3] for proof):
Corollary 2.3. Let $F$ be any field. For a quaternion algebra $Q$ over $F$, let $K = F(Q)$. For any field $E \supseteq F$ with $[E : F] < \infty$, $E$ splits $Q$ if and only if $E \supseteq \mathcal{V}_P$ for some $P \in \mathcal{P}(K/F)$.

Let $F$ be a global field, that is, $F$ is either an algebraic number field (i.e., a finite extension of $\mathbb{Q}$) or an algebraic function field in one variable over a finite field. If $Q$ is a quaternion algebra over $F$, we define the support of $Q$ as follows:

$$\text{supp}(Q) = \{ p \in \mathcal{P}(F) \mid Q \otimes_F \hat{F}_p \text{ is nonsplit} \}.$$ 

Recall that if $p \in \mathcal{P}(F)$ is an archimedean prime spot, then the field $\hat{F}_p$ is isomorphic to $\mathbb{R}$ or $\mathbb{C}$. If $p$ is non-archimedean, then $\hat{F}_p$ is a local field, i.e., a field with complete discrete valuation and with finite residue field.

Lemma 2.4 below can be easily shown (or see [Ha, Remark 4.4]).

Lemma 2.4. Let $Q$ and $Q'$ be quaternion algebras over a global field $F$, and let $E$ be a finite degree field extension of $F$. If $\text{supp}(Q') \subseteq \text{supp}(Q)$, then $\text{supp}(Q' \otimes_F E) \subseteq \text{supp}(Q \otimes_F E)$.

The next two theorems are well-known and are of fundamental importance.

Hasse Principle (special case) 2.5. For a global field $F$, let $Q$ and $Q'$ be quaternion algebras over $F$. Then $Q \cong Q'$ if and only if $Q \otimes_F \hat{F}_p \cong Q' \otimes_F \hat{F}_p$ for all $p \in \mathcal{P}(F)$. In particular, $Q$ is split if and only if $\text{supp}(Q)$ is the empty set.

Hilbert’s Reciprocity Law 2.6. Let $F$ be a global field. For a quaternion algebra $Q$ over $F$, the set $\text{supp}(Q)$ is finite with even cardinality. Further, given any finite subset $S$ of $\mathcal{P}(F)$ with $|S|$ even, there is a unique quaternion algebra $Q$ over $F$ with $\text{supp}(Q) = S$.

If $Q$ is split over a global field $F$, all the assertions we will make are vacuously true. Thus we will exclude this trivial case.

Proposition 2.7. Let $F$ be any field. Suppose that $Q$ is a quaternion division algebra over $F$. Let $K = F(Q)$. For $r \in F^* - F^{*2}$, the following are equivalent:

(i) $F(\sqrt{r})$ splits $Q$.
(ii) $F(\sqrt{r})$ is the residue field of a place of $K/F$.

Further, if $F$ is a global field and $\text{supp}(Q) = \{ p_1, p_2, \ldots, p_n \}$, then (i) and (ii) are also equivalent to:

(iii) $r \notin \bigcup_{p \in \mathcal{P}(F)} \hat{F}_p^{*2}$.

Proof.

(i) $\Rightarrow$ (ii) Since $Q \otimes_F F(\sqrt{r})$ is split over $F(\sqrt{r})$, $K(\sqrt{r})$ is a rational function field over $F(\sqrt{r})$ by Lemma 2.2. Thus we can take a place $P'$ of
\( K(\sqrt{r})/F(\sqrt{r}) \) of degree \( P'(r) = 1 \) and a place \( P \) of \( K/F \) with \( P'|P \). For the respective residue fields of \( P' \) and \( P \), we have \( F(\sqrt{r}) = \overline{P'} \supseteq \overline{P} \supseteq F \). However, since \( Q \) is not split over \( F \), there exists no place \( P \) of degree \( P = 1 \) by Lemma 2.22 again and so \( \overline{P} \neq F \). Since \( [F(\sqrt{r}) : F] = 2 \), we have \( \overline{P'} = F(\sqrt{r}) \).

(ii) \( \Rightarrow \) (i) This is immediate as a consequence of Corollary 2.3.

(i) \( \Rightarrow \) (iii) For each \( \mathfrak{p} \in \text{supp}(Q) \), \( Q \otimes_F \hat{F}_\mathfrak{p} \) is nonsplit. However, \( Q \otimes_F \hat{F}_\mathfrak{p}(\sqrt{r}) \) is split since \( Q \otimes_F F(\sqrt{r}) \) is split. Hence \( \hat{F}_\mathfrak{p}(\sqrt{r}) \neq \hat{F}_\mathfrak{p} \). In other words, \( r \notin \hat{F}_\mathfrak{p} \cup \hat{F}_\mathfrak{p} \cup \cdots \cup \hat{F}_\mathfrak{p} \).

(iii) \( \Rightarrow \) (i) By the Hasse Principle 2.5, it suffices to show that \( Q \otimes_F \overline{F(\sqrt{r})}_{\mathfrak{p}} \) is split for every prime spot \( \mathfrak{p} \) of \( F(\sqrt{r}) \). To see this, assume that \( \mathfrak{p} \) is the restriction of \( \mathfrak{p} \) to \( F \). We have two possibilities for this \( \mathfrak{p} \): If \( \mathfrak{p} \notin \text{supp}(Q) \), then \( Q \otimes_F \overline{F(\sqrt{r})}_{\mathfrak{p}} \) is clearly split since \( Q \otimes_F \hat{F}_\mathfrak{p} \) is already split.

On the other hand, if \( \mathfrak{p} \in \text{supp}(Q) \), then \( Q \otimes_F \hat{F}_\mathfrak{p} \) is nonsplit. However, by (iii) \( \hat{F}_\mathfrak{p}(\sqrt{r}) = F(\sqrt{r})_{\mathfrak{p}} \) is a quadratic extension of \( \hat{F}_\mathfrak{p} \). Now, recall the well-known results that any local field \( L \) or \( L = \mathbb{R} \) has a unique nonsplit quaternion algebra, which is split by each quadratic extension of \( L \). (For this, use e.g., [Re, Theorems 31.4, 31.8, and 31.9] and the facts that \( \text{Br}(\mathbb{R}) \cong \mathbb{Z}/2\mathbb{Z} \) and \( \text{Br}(\mathbb{C}) \cong 0 \).) Thus, \( Q \otimes_F \hat{F}_\mathfrak{p}(\sqrt{r}) \) must be split. Hence, \( Q \otimes_F \hat{F}_\mathfrak{p}(\sqrt{r}) = Q \otimes_F F(\sqrt{r})_{\mathfrak{p}} \) is split for every prime spot \( \mathfrak{p} \) of \( F(\sqrt{r}) \).  

\[ \square \]

### 3. Computation of \( \bigcap_{P \in \mathfrak{p}(K/F)} \text{Br}(K\overline{V}_P/K) \).

Let \( F \) be a global field and let \( Q = (a, b/F) \) be a quaternion division algebra over \( F \) with \( \text{supp}(Q) = \{ \mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_n \} \). Recall that \( n \) is finite and even by Hilbert’s Reciprocity Law 2.6. Let \( K = F(Q) = F(x, \sqrt{ax^2 + b}) \). Consider the map

\[ (2) \quad h : \text{Br}(K) \longrightarrow \prod_{P \in \mathfrak{p}(K/F)} \text{Br}(K\overline{V}_P). \]

The purpose of this section is to explicitly compute

\[ \ker(h) = \bigcap_{P \in \mathfrak{p}(K/F)} \text{Br}(K\overline{V}_P/K) \]

(see Theorem 3.7) which will be used to prove our main theorems in Chapter 4. For this, we will first compute \( \bigcap_{P \in \mathfrak{p}(K/F)} \text{Br}(\overline{V}_P/F) \) by utilizing the local-global principle. This intersection can be directly described, in Proposition 3.2, in terms of the quaternion algebra \( Q \).
To begin with, we want to give explicit calculation of \( \bigcap_{\deg(P)=2} \text{Br}(\mathcal{V}_P/F) \) over all the places \( P \in \mathbb{P}(K/F) \) with \( \deg(P)=2 \). For \( \text{supp}(Q) = \{ p_1, p_2, \ldots, p_n \} \), we define

\[
\mathcal{F}_Q = \{ r \in F \mid r \notin \hat{F}_{p_1}^{*2} \cup \hat{F}_{p_2}^{*2} \cup \cdots \cup \hat{F}_{p_n}^{*2} \}.
\]

Proposition 2.7 yields another description of \( \mathcal{F}_Q \):

\[
\mathcal{F}_Q = \{ r \in F^* - F^{*2} \mid F(\sqrt{r}) \text{ is the residue field of a place of } K/F \}.
\]

Therefore, we have

\[
\bigcap_{\deg(P)=2} \text{Br}(\mathcal{V}_P/F) = \bigcap_{r \in \mathcal{F}_Q} \text{Br}(F(\sqrt{r})/F).
\]

We will use a number of times below the following well-known combinatorial fact:

**Lemma 3.1.** Let \( T \) be a set with \( |T| = n \) where \( n \in \mathbb{N} \). Then the number of subsets with an even number of elements of \( T \) is \( 2^n - 1 \).

We next define

\[
\mathcal{I}_Q = \left\{ [Q'] \mid Q' \text{ is a quaternion algebra over } F \text{ with } \text{supp}(Q') \subseteq \text{supp}(Q) \right\} \subseteq \text{Br}(F).
\]

**Proposition 3.2.** Let \( F \) be a global field. Suppose that \( Q \) is a quaternion division algebra over \( F \). Let \( K = F(Q) \). For the \( \mathcal{F}_Q \) in (3) and \( \mathcal{I}_Q \) in (5), we have

\[
\bigcap_{P \in \mathbb{P}(K/F)} \text{Br}(\mathcal{V}_P/F) = \bigcap_{\deg(P)=2} \text{Br}(\mathcal{V}_P/F) = \bigcap_{r \in \mathcal{F}_Q} \text{Br}(F(\sqrt{r})/F) = \mathcal{I}_Q.
\]

The cardinality of this set is \( 2^{n-1} \) where \( n = |\text{supp}(Q)| \).

**Proof.** We will prove the last equality in (6) first. Any element in the relative Brauer group of \( F(\sqrt{r})/F \) is the class of a quaternion algebra over \( F \) (cf. [Dr, Corollary 1, p. 79]). Thus, any element in the intersection is the class of a quaternion algebra over \( F \). If \( Q' \) is a nonsplit quaternion algebra such that \( \text{supp}(Q') \subseteq \text{supp}(Q) \), then \( \mathcal{F}_Q \subseteq \mathcal{F}_{Q'} \). Thus, for each \( r \in \mathcal{F}_Q \), \( F(\sqrt{r}) \) splits \( Q' \) by Proposition 2.7 applied to \( Q' \). Hence, \( [Q'] \in \bigcap_{r \in \mathcal{F}_Q} \text{Br}(F(\sqrt{r})/F) \). On the other hand, if \( \text{supp}(Q') \nsubseteq \text{supp}(Q) \), we can take \( p \in \text{supp}(Q') - \text{supp}(Q) \). Assume that \( \text{supp}(Q) = \{ p_1, p_2, \ldots, p_n \} \). By the Weak Approximation Theorem (cf. [Ws, p. 8]), there exists \( r \in F \) such that

\[
r \in \hat{F}_{p}^{*2} \text{ but } r \notin \hat{F}_{p_1}^{*2} \cup \hat{F}_{p_2}^{*2} \cup \cdots \cup \hat{F}_{p_n}^{*2}
\]

(for this, recall that if two elements are \( q \)-adically close enough, then they lie in the same \( q \)-adic square class). In other words, \( r \in \mathcal{F}_Q \) but \( r \notin \mathcal{F}_{Q'} \). Thus,
[\mathcal{Q}'] \not\in \text{Br}(\sqrt{r}/F)$ which clearly implies that $[\mathcal{Q}'] \not\in \bigcap_{r \in \mathcal{F}_Q} \text{Br}(\sqrt{r}/F)$.

This assertion gives the last equality in (6).

The middle equality in (6) was already given in (4). Now, we show that
\[
\bigcap_{P \in \mathcal{P}(K/F)} \text{Br}(\mathcal{V}_P/F) = \bigcap_{r \in \mathcal{F}_Q} \text{Br}(\sqrt{r}/F).
\]

One inclusion ($\subseteq$) is clear from (4). For the other inclusion ($\supseteq$), take $[\mathcal{Q}'] \in \bigcap_{r \in \mathcal{F}_Q} \text{Br}(\sqrt{r}/F)$. Then, we may assume that $\mathcal{Q}'$ is a quaternion algebra over $F$ with $\text{supp}(\mathcal{Q}') \subseteq \text{supp}(\mathcal{Q})$ by the last equality of (6). If we put $\mathcal{V} = \mathcal{V}_P$ for $P \in \mathcal{P}(K/F)$, then $[\mathcal{V} : F] < \infty$ and $\mathcal{V}$ splits $\mathcal{Q}$ (as seen in the proof (ii) $\Rightarrow$ (i) of Proposition 2.7). By Lemma 2.4,
\[
\text{supp}(\mathcal{Q}' \otimes_F \mathcal{V}) \subseteq \text{supp}(\mathcal{Q} \otimes_F \mathcal{V}) = \emptyset.
\]

Hence, $\mathcal{V}$ splits $\mathcal{Q}'$.

Finally, Hilbert’s Reciprocity Law 2.6 implies that the number of elements in $\mathcal{I}_Q$ equals the number of subsets with an even number of elements of a set with $n$ elements. By Lemma 3.1, the cardinality of this set is $2^{n-1}$. □

Example 3.3. Let $\mathcal{Q} = (-1, -1/\mathbb{Q})$ be a quaternion algebra over $\mathbb{Q}$ and let $K = \mathbb{Q}(\mathcal{Q}) = \mathbb{Q}(x, \sqrt{-x^2 - 1})$. It is easy to see that $\text{supp}(\mathcal{Q}) = \{2, \infty\}$ where 2 is the dyadic spot and $\infty$ is the real infinite spot of $\mathbb{Q}$. Thus, we have
\[
\bigcap_{P \in \mathcal{P}(K/F)} \text{Br}(\mathcal{V}_P/F) = \{0, [\mathcal{Q}]\}.
\]

In this example, we can explicitly describe all the quadratic residue fields (see (iii) and (vi) below). For $a_1, \cdots, a_n \in F^*$, let $(a_1, \cdots, a_n)$ be the diagonal quadratic form $a_1x_1^2 + \cdots + a_nx_n^2$. Assume that $r$ is a square-free integer. Then the following are equivalent:

(i) $\mathbb{Q}(\sqrt{r})$ is a residue field of a place of $K/\mathbb{Q}$.
(ii) $r \in \mathcal{I}_Q$, i.e., $r \not\in \widehat{\mathbb{Q}}_2^2 \cup \mathbb{R}^2$.
(iii) $r < 0$ and $r \equiv 1 \pmod{8}$.
(iv) $r < 0$ and the quadratic form $(1, 1, 1, r)$ is isotropic over $\widehat{\mathbb{Q}}_2$.
(v) $r < 0$ and the quadratic form $(1, 1, 1, r)$ is isotropic over $\mathbb{Q}$.
(vi) $r < 0$ and $-r$ is the sum of three squares in $\mathbb{Q}$.

Indeed, (i) $\iff$ (ii) is given by Proposition 2.7 since $\text{supp}(\mathcal{Q}) = \{2, \infty\}$. For (ii) $\iff$ (iii), clearly $r \in \mathbb{R}^2 \iff r > 0$. The fact that $r \in \widehat{\mathbb{Q}}_2^2 \iff r \equiv 1 \pmod{8}$ is well-known (see [La, Corollary 2.24, p. 162]). For (ii) $\iff$ (iv), recall (cf. [OM, 63:17, p. 169]) that over a local field, a 4-dimensional anisotropic quadratic form has determinant 1 modulo squares. Because $\text{supp}(\mathcal{Q}) = \{2, \infty\}$, the quadratic form $(1, 1, 1, 1)$ is anisotropic over $\widehat{\mathbb{Q}}_2$ as it is the norm form of $(-1, -1/\widehat{\mathbb{Q}}_2)$. To show (iv) $\iff$ (v), $(1, 1, 1, r)$ is certainly isotropic over $\mathbb{R}$ and
for any prime spot \( p \) corresponding to an odd prime \( p \), observe that \( \langle 1, 1, 1 \rangle \)
is already isotropic over \( \mathbb{Q}_p \) and so is \( \langle 1, 1, 1, r \rangle \). Then, this equivalence is an immediate consequence of the Hasse-Minkowski Theorem (cf. [La, p. 168]). (v) \( \Leftrightarrow \) (vi) is obvious.

For two central simple algebras \( A \) and \( B \) over \( F \), we will write \( A \sim B \) if \( [A] = [B] \) in \( \text{Br}(F) \). The following examples will be used later in the sequel.

**Example 3.4.** Let \( n > 0 \) be an odd integer. Suppose that \( p_1, p_2, \ldots, p_n \)
are distinct odd prime numbers such that each \( p_i \equiv 3 \pmod{4} \). Let \( Q = (-1, m/Q) \) where \( m = p_1 p_2 \cdots p_n \) and let \( K = \mathbb{Q}(Q) = \mathbb{Q}(x, \sqrt{-x^2 + m}) \). Then

\[
\text{supp}(Q) = \{2, p_1, \ldots, p_n\},
\]

where \( 2 \) is the dyadic spot and each \( p_i \) denotes the prime spot corresponding to the odd prime \( p_i \). To see this, let \( P_i = (-1, p_i/Q) \). For a fixed \( i \), we observe that \( \mathbb{R} \) splits \( P_i \) since \( p_i > 0 \) and that \( \mathbb{Q}_p \) splits \( P_i \) for \( p \) any odd prime different from \( p_i \) since \( -1 \) and \( p_i \) are both \( p \)-adic units. However, \( P_i \otimes_{\mathbb{Q}} \mathbb{Q}_p \) is not split since \( -1 \) is not a square in \( \mathbb{Z}/p_i \mathbb{Z} \) as \( p_i \equiv 3 \pmod{4} \). By Hilbert’s Reciprocity Law 2.6, we have \( \text{supp}(P_i) = \{2, p_i\} \). Since

\[
Q \sim P_1 \otimes_{\mathbb{Q}} P_2 \otimes_{\mathbb{Q}} \cdots \otimes_{\mathbb{Q}} P_n
\]

and \( n \) is odd, it follows that \( \text{supp}(Q) = \{2, p_1, \ldots, p_n\} \) as claimed. Then, Proposition 3.2 yields

\[
\bigcap_{P \in \mathcal{P}(K/Q)} \text{Br}(\mathcal{V}_P/Q) = \{0, [P_1], [P_1 \otimes_{\mathbb{Q}} P_2], \ldots, [P_1 \otimes_{\mathbb{Q}} \cdots \otimes_{\mathbb{Q}} P_{n-1}], [Q]\},
\]

where the \( i_j \) range over all distinct numbers in \( \{1, 2, \ldots, n\} \), and the cardinality of this set is \( 2^{\text{supp}(Q)}-1 = 2^n \). In particular, if \( n = 1 \) (so \( m = p_1 \)), then

\[
\bigcap_{P \in \mathcal{P}(K/Q)} \text{Br}(\mathcal{V}_P/Q) = \{0, [Q]\}.
\]

**Example 3.5.** Let \( n > 0 \) be an even integer. Suppose that \( p_1, p_2, \ldots, p_n \)
are as in Example 3.4. Let \( Q = (-1, m/Q) \) where \( m = p_1 p_2 \cdots p_n \) and let \( K = \mathbb{Q}(Q) = \mathbb{Q}(x, \sqrt{-x^2 + m}) \). If we let \( P_i = (-1, p_i/Q) \), then we have

\[
Q \sim P_1 \otimes_{\mathbb{Q}} P_2 \otimes_{\mathbb{Q}} \cdots \otimes_{\mathbb{Q}} P_n.
\]

Applying the similar arguments in Example 3.4, we have

\[
\text{supp}(Q) = \{p_1, p_2, \ldots, p_n\}
\]
since \( n \) is even. Then, Proposition 3.2 yields

\[
\bigcap_{P \in \mathcal{P}(K/Q)} \text{Br}(\mathcal{V}_P/Q) = \{0, [P_1 \otimes_{\mathbb{Q}} P_2], [P_1 \otimes_{\mathbb{Q}} \cdots \otimes_{\mathbb{Q}} P_4], \ldots, [P_1 \otimes_{\mathbb{Q}} \cdots \otimes_{\mathbb{Q}} P_{n-2}], [Q]\},
\]
where the $i_j$ range over all distinct numbers in $\{1, 2, \ldots, n\}$, and the cardinality of this set is $2^{n-1}$. In particular, if $n = 2$ (so $m = p_1 p_2$), then

$$\bigcap_{P \in \mathbb{P}(K/Q)} \text{Br}(\mathbb{V}_P/Q) = \{0, [Q]\}. $$

Let $Q = (a, b/F)$ and $K = F(Q) = F(x, \sqrt{ax^2 + b})$. Fix $r \in Fr$. For $P \in \mathbb{P}(K/F)$ and $P' \in \mathbb{P}(K(\sqrt{r})/F(\sqrt{r}))$, denote by $G_P$ and $G_{P'}$ the absolute Galois groups of the residue fields $\mathbb{V}_P$ and $\mathbb{V}_{P'}$, respectively. Note that $G'$ depends on $r \in Fr$. If $P'$ is a place above $P$, note that $G_{P'}$ is a subgroup of $G_P$ and $|G_P : G_{P'}| = [\mathbb{V}_{P'} : \mathbb{V}_P] = 1$ or $2$.

**Lemma 3.6.** Keeping the notation as above, let $\chi_P \in X(G_P)$, the character group of $G_P$. Suppose that $\chi_P|_{G_{P'}} = 0$ for all $P' \in \mathbb{P}(K(\sqrt{r})/F(\sqrt{r}))$ with $P'|P$ for all $r \in Fr$. Then $\chi_P = 0$.

**Proof.** Consider the map $\varphi : F^*/F^{*2} \rightarrow \mathbb{V}_P/F^{*2}$. By Kummer theory, we have

$$|\ker(\varphi)| = [F\{\sqrt{a} \in \ker(\varphi)\}: F] \leq [\mathbb{V}_P : F] < \infty.$$ 

Now, we show that $FrF^{*2}/F^{*2}$ is infinite. To see this, let

$$A = \{p \in P(F) \mid p \text{ is finite and } p \not\in \text{supp}(Q)\} $$

and let $B$ be any finite subset of $A$. Note that for any $p \in B$, by the Weak Approximation Theorem (cf. [Ws, p. 8]), there is $r \in Fr$ with $v_p(r)$ odd. From the surjective map

$$\psi : F^*/F^{*2} \rightarrow \bigoplus_{p \in B} \mathbb{Z}/2\mathbb{Z},$$

given by $aF^{*2} \mapsto (\cdots, v_p(a) + 2\mathbb{Z}, \cdots)$, we have $\psi(FrF^{*2}/F^{*2}) \geq 2^{[B]}$. Since this is true for any finite subset $B$ of $A$, it follows that $\psi(FrF^{*2}/F^{*2})$ and thus $FrF^{*2}/F^{*2}$ is infinite. This implies that $\varphi(FrF^{*2}/F^{*2})$ is infinite, since $\ker(\varphi) < \infty$. Thus, it is possible to take $r_1, r_2 \in Fr$ such that $\mathbb{V}_P(\sqrt{r_1}) \neq \mathbb{V}_P(\sqrt{r_2})$. For $i = 1$ or $2$, suppose that $P_i \in \mathbb{P}(K(\sqrt{r_i})/F(\sqrt{r_i}))$ with $\deg(P_i) = 1$. If $r_i \not\in \mathbb{V}_P$ for some $i$, then $G_{P_i} = G_P$ and so $\chi_P = \chi_{P_i} = 0$, as desired. Thus, we assume that $r_1, r_2 \not\in \mathbb{V}_P$. Clearly, $\mathbb{V}_{P_1} = \mathbb{V}_{P_2}$. If we let $G_{P_1} = \text{Gal}(F_{sep}/\mathbb{V}_{P_1})$, then

$$|G_{P_1} : G_{P_2}| = |G_{P_1} : G_{P_2}| = 2 \text{ and } G_{P_1} \neq G_{P_2}.$$ 

Since $G_P = G_{P_1} \cdot G_{P_2}$, we have $\chi_{P_1} = \chi_{P_2} = 0$. \qed

Before we discuss the main theorem of this section, it will be convenient to define some new terminology. Let $K$ be an algebraic function field in one variable (of genus 0) over a constant field $F$. For the scalar extension
map $\alpha : \Br(F) \to \Br(K)$, a class $[B] \in \Br(K)$ is called a constant class if $[B] \in \im(\alpha)$.

For the map $h$ in (2) above, we compute the kernel of $h$. The elements in $\ker(h)$ turn out to be all constant classes of quaternion algebras over $F$. As in Proposition 3.2, it suffices to take the intersection over all the quadratic residue fields.

**Theorem 3.7.** Let $F$ be a global field. Suppose that $Q$ is a quaternion division algebra over $F$. Let $K = F(Q)$. For the map $h$ in (2), the $\mathcal{F}_Q$ in (3), and $\mathcal{I}_Q$ in (5), we have

$$\ker(h) = \bigcap_{P \in \mathcal{P}(K/F)} \Br(K \mathcal{V}_P/K) = \bigcap_{\deg(P)=2} \Br(K(\sqrt{r})/K) = \{ [Q' \otimes_F K] \mid [Q'] \in \mathcal{I}_Q \}.$$  

The cardinality of this set is $2^{n-2}$ where $n = |\text{supp}(Q)|$.

**Proof.** Let us show the last equality of (7) first. For each $r \in \mathcal{F}_Q$, we have the following diagram with exact rows:

$$0 \longrightarrow \mathbb{Z}_2 \longrightarrow \Br(F) \overset{\text{res}}{\longrightarrow} \Br(F_{\text{sep}} \cdot K/K) \overset{\beta}{\longrightarrow} \bigoplus_{P \in \mathcal{P}(K/F)} X(G_P)$$

$$\overset{\text{res}}{\longrightarrow} \Br(F(\sqrt{r})) \overset{\gamma}{\longrightarrow} \Br(F_{\text{sep}} \cdot K(\sqrt{r})/K(\sqrt{r})) \overset{\delta}{\longrightarrow} \bigoplus_{P \mid P'} X(G_{P'}).$$

Here the top sequence is the sequence (1) and the bottom sequence comes also from (1) since $F(\sqrt{r})$ splits $Q$ (or, from the Auslander-Brumer-Fadeev Theorem (cf. [AB], [Fa] or [FS])). Each component $e_p \cdot \text{res}$ in the right vertical map in (8) is the ramification index $e_p = e(v_P/r_P)$ times the natural restriction map from $X(G_P)$ to $X(G_P')$. The right square in (8) is commutative (see [Sa, Theorem 10.4]). The left square is clearly commutative, since all the maps are restriction maps. Note that for each $P \in \mathcal{P}(K/F)$, $e_p = 1$, since $r \in F$ and so $v_P(r) = 0$ (cf. [St, Theorem III.6.3, p. 103]).

Let $[B] \in \bigcap_{r \in \mathcal{F}_Q} \Br(K(\sqrt{r})/K)$. Since $\exp(B) \leq 2$ and $\text{char}(F) \neq 2$, $[B]$ is split by $F_{\text{sep}} \cdot K$. Obviously, for any $r \in \mathcal{F}_Q$, $\beta([B]) \in \ker(\oplus e_p \cdot \text{res})$ from the commutativity of the right square. Put

$$\beta([B]) = \sum_{\text{finite}} \chi_P.$$  

Applying Lemma 3.6 to each $\chi_P$, we have $\beta([B]) = 0$ and thus $[B] \in \im(\alpha)$ by the exactness at $\Br(F_{\text{sep}} \cdot K/K)$. Suppose that $[A]$ is a preimage of $[B]$ in $\Br(F)$. Since $\gamma$ is injective, we have $[A] \in \Br(F(\sqrt{r})/F)$ for each $r \in \mathcal{F}_Q$ by
the commutativity of the left square in (8). Thus \([B]\) is of the form \([Q' \otimes_F K]\)
where
\[
[Q'] \in \bigcap_{r \in F_Q} \text{Br}(F(\sqrt[r]{\tau})/F) = \mathcal{I}_Q
\]
by Proposition 3.2. Conversely, for each \([Q'] \in \mathcal{I}_Q\) we clearly have \([Q' \otimes_F K] \in \bigcap_{r \in F_Q} \text{Br}(K(\sqrt[r]{\tau})/K)\).

Moreover, by Proposition 3.2, there are \(2^{n-1}\) elements in \(\mathcal{I}_Q\). Since \(\text{Br}(K/F) = \{0, [Q]\}\) from Proposition 2.1, two different elements \([Q']\) and \([Q' \otimes_F Q]\) in \(\text{Br}(F)\) have the same image \([Q' \otimes_F K]\) in \(\text{Br}(F_{\text{sep}} K/K)\). This observation shows that the set in (7) contains \(2^{n-2}\) elements of the form \([Q' \otimes_F K]\).

The first equality of (7) is clear from the definition of the map \(h\), and the third one is also clear by Proposition 2.7.

Finally, we verify the second equality of (7). One inclusion (\(\subseteq\)) is clear. In order to show the other inclusion (\(\supseteq\)), we use the fact just proved that
\[
\bigcap_{\deg(P)=2} \text{Br}(K\sqrt{\Gamma}/K) = \{[Q' \otimes_F K] | [Q'] \in \mathcal{I}_Q\}.
\]
For each \([Q'] \in \mathcal{I}_Q\) and \(P \in \mathbb{P}(K/F)\), we apply Proposition 3.2 to have \([Q'] \in \text{Br}(\sqrt{\Gamma}/F)\) and therefore \([Q' \otimes_F K] \in \text{Br}(K\sqrt{\Gamma}/K)\). This completes the proof.

**Example 3.8.** For an odd integer \(n > 0\), let \(p_1, p_2, \ldots, p_n\) be distinct odd prime numbers with each \(p_i \equiv 3 \pmod{4}\). Let \(Q = (-1, m/Q)\) where \(m = p_1 p_2 \ldots p_n\) and let \(K = \mathbb{Q}(Q) = \mathbb{Q}(x, \sqrt{-x^2 + m})\) as in Example 3.4. Recall that \(\text{supp}(Q) = \{2, p_1, \ldots, p_n\}\). For each \(P_i = (-1, p_i/Q)\), let
\[
B_i = P_i \otimes_Q K \cong (-1, p_i/K).
\]
If we set \(k = \frac{n-1}{2} \in \mathbb{Z}\), then Theorem 3.7 and Example 3.4 yield
\[
\bigcap_{P \in \mathbb{P}(K/Q)} \text{Br}(K\sqrt{\Gamma}/K) = \{0, [B_{i_1}], [B_{i_1} \otimes_K B_{i_2}], \ldots, [B_{i_1} \otimes_K \cdots \otimes_K B_{i_k}]\},
\]
where the \(i_j\) range over all distinct numbers in \(\{1, 2, \ldots, n\}\), and the cardinality of this set is \(2^{\text{supp}(Q)-2} = 2^{n-1}\). In particular, if \(n = 1\) (so \(m = p_1\)), then we have
\[
\bigcap_{P \in \mathbb{P}(K/Q)} \text{Br}(K\sqrt{\Gamma}/K) = \{0\}.
\]

**Example 3.9.** For an even integer \(n > 0\), let \(p_1, p_2, \ldots, p_n\) be as in Example 3.5. Let \(Q = (-1, m/Q)\) where \(m = p_1 p_2 \ldots p_n\) and let \(K = \mathbb{Q}(Q) = \mathbb{Q}(x, \sqrt{-x^2 + m})\). Recall that \(\text{supp}(Q) = \{p_1, p_2, \ldots, p_n\}\). For each
\( P_i = (-1, p_i/\mathbb{Q}) \), let \( B_i = P_i \otimes_{\mathbb{Q}} K \cong (-1, p_i/K) \). If we set \( k = \frac{n}{2} \in \mathbb{Z} \), then Theorem 3.7 and Example 3.5 yield

\[
\bigcap_{P \in \mathfrak{p}(K/\mathbb{Q})} \text{Br}(K\mathfrak{V}_P/K) = \{0, [B_{i_1} \otimes_K B_{i_2}], [B_{i_1} \otimes_K \cdots \otimes_K B_{i_k}], \ldots, [B_{i_1} \otimes_K \cdots \otimes_K B_{i_k}]\},
\]

where the \( i_j \) range over all distinct numbers in \( \{1, 2, \ldots, n\} \), and the cardinality of this set is \( 2^{n-2} \). In particular, if \( n = 2 \) (so \( m = p_1 p_2 \)), then we have

\[
\bigcap_{P \in \mathfrak{p}(K/\mathbb{Q})} \text{Br}(K\mathfrak{V}_P/K) = \{0\}.
\]

4. Obstruction to Hasse principle for the Brauer group of a function field of genus 0.

The fundamental and profound result on the Brauer group of a global field \( F \) provides an exact sequence (cf. [We, Theorem 2, p. 206, and Theorem 4, p. 164])

\[
0 \longrightarrow \text{Br}(F) \overset{i}{\longrightarrow} \bigoplus_{p \in P(F)} \text{Br}(\hat{F}_p) \overset{\text{inv}}{\longrightarrow} \mathbb{Q}/\mathbb{Z} \longrightarrow 0, \tag{9}
\]

where the map \( i \) is the direct sum of scalar extension maps \( \text{Br}(F) \to \text{Br}(\hat{F}_p) \) for \( p \in P(F) \) and the map \( \text{inv} \) is the invariant map, computed locally on each component \( p \), that is, \( \text{inv} = \oplus_p \text{inv}_{\hat{F}_p} \). In particular, the injectivity of the natural map \( i \) in (9) asserts the classical Hasse principle for \( \text{Br}(F) \) (cf. (2.5) for the case of quaternion algebras). In this section, we consider the analogous possible Hasse principles for the Brauer groups of algebraic function fields in one variable of genus 0 over global fields.

We begin with the case of rational function fields. For a field \( k \), let \( k(x) \) be the rational function field over \( k \).

**Lemma 4.1.** Let \( F \) be a global field and \( F_{\text{sep}} \) the separable closure of \( F \). Then the map

\[
j : \text{Br}(F_{\text{sep}}(x)/F(x)) \longrightarrow \prod_{p \in P(F)} \text{Br}(\hat{F}_p(x)) \tag{10}
\]

is injective.
Proof. We have the following commutative diagram with exact rows from the Auslander-Brumer-Faddeev Theorem (cf. [AB], [Fa] or [FS]):

\[
\begin{array}{cccc}
0 & \longrightarrow & \text{Br}(F) & \longrightarrow & \text{Br}(F_{\text{sep}}(x)/F(x)) \longrightarrow & \bigoplus_{P \in \mathcal{P}(F(x)/F)} X(G_P) \\
& & i & \downarrow & j_0 & \downarrow \psi \\
0 & \longrightarrow & \prod_{p \in P(F)} \text{Br}(\widehat{F}_p) & \longrightarrow & \prod_{p \in P(F)} \text{Br}(\widehat{F}_p(\text{sep})(x)/\widehat{F}_p(x)) & \longrightarrow & \prod_{p \in P(F) \mid P'} \bigoplus_{P'} X(G_{P'}),
\end{array}
\]  

where \( P' \in \mathcal{P}(\widehat{F}_p(\text{sep})/\widehat{F}_p) \) is a place above \( P \).

We claim that the map \( \psi \) in (11) is injective. For this, take a nontrivial character \( \chi \in X(G_P) \). Observe first that \( \chi(G_P) \subseteq \mathbb{Q}/\mathbb{Z} \) is finite and so cyclic. If \( M \subseteq F_{\text{sep}} \) is the fixed field of \( \ker(\chi) \), then \( M \) is a cyclic Galois extension of \( \overline{V}_P =: E \) and the order of \( \chi \) equals the degree \( [M : E] \). For \( p \in P(F) \), take \( P' \in \mathcal{P}(\widehat{F}_p(x)/\widehat{F}_p) \) with \( P'|P \). Then for the image \( \chi' \) of \( \chi \) in \( X(G_{P'}) \), the order of \( \chi' \) equals the degree \( [M_{\mathcal{P}'} : \overline{V}_{\mathcal{P}'}/\overline{V}_P] \). By Tchebotarev density (cf. [FJ, Theorem 5.6]), there exists \( \mathfrak{p} \in \mathcal{P}(E) \) such that for every \( \mathfrak{p}' \in \mathcal{P}(M) \) with \( \mathfrak{p}'|\mathfrak{p} \), we have \( [M : E] = [M_{\mathcal{P}'} : \overline{V}_{\mathcal{P}'}/\overline{V}_P] \). Choose \( p \in P(F) \) as the restriction of \( \mathfrak{p}' \) to \( F \); then there is \( P' \in \mathcal{P}(\widehat{F}_p(x)/\widehat{F}_p) \) with \( P'|P \) and \( \overline{V}_{\mathcal{P}'} = \overline{V}_p \). Then we have \( M_{\mathcal{P}'} = \overline{M}_{\mathcal{P}'} \) and therefore the order of \( \chi' \) equals that of \( \chi \). Hence, the map \( \psi \) is injective as claimed.

Now, since the map \( i \) in (11) is also injective by (9), we conclude that the map \( j_0 \) in (11) and so the map \( j \) in (10) is injective.

\[ \Box \]

For a field \( k \), let \( \text{Br}(k)' \) denote the subgroup of \( \text{Br}(k) \) consisting of all elements of order relatively prime to \( p \) if \( \text{char}(k) = p > 0 \). If \( \text{char}(k) = 0 \), set \( \text{Br}(k)' = \text{Br}(k) \). Notice then that

\[ \text{Br}(F_{\text{sep}}(x)/F(x))' = \text{Br}(F(x))' \]

since \( \text{Br}(F_{\text{sep}}(x))' = 0 \) (see [FS, Lemma 2, p. 51]). Thus, we have:

**Corollary 4.2.** Let \( F \) be a global field and let \( F(x) \) be the rational function field over \( F \). Then, the map

\[ j : \text{Br}(F(x))' \longrightarrow \prod_{p \in P(F)} \text{Br}(\widehat{F}_p(x)) \]

is injective.

**Remark 4.3.** Corollary 4.2 can be also proved by specialization argument using the fact that every global field is Brauer-Hilbertian in the sense of Fein, Saltman, and Schacher. For details on Brauer-Hilbertian fields, see [FSS].
Let $Q = (a, b/F)$ be a quaternion division algebra over a global field $F$ and let $K = F(Q) = F(x, \sqrt{ax^2 + b})$. We want to describe the kernel of the map
\begin{equation}
(13) \quad h_1 : \text{Br}(K) \to \prod_{p \in \mathcal{P}(F)} \text{Br}(\widehat{F}_p(Q)).
\end{equation}

**Theorem 4.4.** Let $F$ be a global field and $K = F(Q)$ as above. For the map $h_1$ in (13) and the set $\mathcal{I}_Q$ in (5), we have
\[ \ker(h_1) = \{ [Q'] \otimes_F K | [Q'] \in \mathcal{I}_Q \}. \]

**Proof.** For each $p \in \mathcal{P}(K/F)$ and $Q' \in \mathcal{I}_Q$, we first show that $Q' \otimes_F \widehat{F}_p(Q)$ is split. For this, it suffices to consider $p \in \text{supp}(Q')$. Then, $Q' \otimes_F \widehat{F}_p$ is nonsplit as is $Q \otimes_F \widehat{F}_p$. It follows that $Q' \otimes_F \widehat{F}_p \cong Q \otimes_F \widehat{F}_p$ since there exists a unique (up to isomorphism) quaternion division algebra over $\widehat{F}_p$. Hence, $\widehat{F}_p(Q)$ splits $Q'$.

We show the other inclusion. Fix $P \in \mathcal{P}(K/F)$ and write the corresponding residue field as $V$, instead of $V_P$ for convenience. For each extension $\mathfrak{p}$ of $p$ to $V$, let $\widehat{V}_{\mathfrak{p}}$ be the completion of $V$ at $\mathfrak{p}$. Then we have the following commutative diagram
\begin{equation}
(14) \quad \begin{array}{ccc}
\text{Br}(F(Q))' & \to & \prod_{p \in \mathcal{P}(F)} \text{Br}(\widehat{F}_p(Q)) \\
\downarrow & & \downarrow \\
\text{Br}(V(Q))' & \to & \prod_{\mathfrak{p} \in \mathcal{P}(V)} \text{Br}(\widehat{V}_{\mathfrak{p}}(Q)).
\end{array}
\end{equation}

Since $V$ splits $Q$ by Corollary 2.3, $V(Q)$ is a rational function field over $V$. By Corollary 4.2, the map $j$ in the diagram (14) is injective. Therefore, any $[B] \in \ker(h_1)$ becomes trivial in $\text{Br}(V(Q))$, which implies that $[B] \in \ker(h)$ by Theorem 3.7. \hfill \square

Let $V$ be a discrete valuation ring with its quotient field $K$. Let $\widehat{K}$ be the completion of $K$ with respect to $V$ and $\widehat{K}_{\text{nr}}$ the maximal unramified extension of $\widehat{K}$. Denote by $X(G_V)$ the character group of the absolute Galois group of the residue field $V$. There is a short exact sequence:
\begin{equation}
(15) \quad 0 \to \text{Br}(V) \to \text{Br}(\widehat{K}_{\text{nr}}/K) \xrightarrow{\text{ram}} X(G_V) \to 0.
\end{equation}

In order to define the ramification map $\text{ram}$, recall that for $[B] \in \text{Br}(\widehat{K}_{\text{nr}}/K)$, $\text{ram}([B])$ is computed by first extending scalars to the completion $\widehat{K}$ and then applying the map
\[ H^2(G_V, \widehat{K}_{\text{nr}}^*) \to H^2(G_V, \mathbb{Z}) \cong X(G_V) \]
induced by the valuation. For details, see [Sa, Theorem 10.3].

Let \( \hat{V} \) be the completion of \( V \), which is a discrete valuation ring with quotient field \( \hat{K} \); note that the residue field of \( \hat{V} \) is isomorphic to \( \hat{V} \) and further we have

\[
\text{Br}(\hat{V}) \cong \text{Br}(\hat{V}) \hookrightarrow \text{Br}(\hat{K}).
\]

See [JW, Theorem 2.8 (b) and Theorem 5.6 (a)] for the isomorphism. These facts will be used in proving Theorem 4.5 below.

Making use of the results in Section 3, we now describe the kernel of the map:

\[
h_2 : \text{Br}(K) \longrightarrow \prod_{P \in \mathbb{P}(K/F)} \text{Br}(\hat{K}_P).
\]

**Theorem 4.5.** Let \( F \) be a global field. Suppose that \( Q = (a, b/F) \) is a quaternion division algebra over \( F \). Let \( K = F(Q) = F(x, \sqrt{ax^2 + b}) \). For the map \( h_2 \) in (17) and the set \( \mathcal{I}_Q \) in (5), we have

\[
\ker(h_2) = \{ [Q' \otimes_F K] \mid [Q'] \in \mathcal{I}_Q \}.
\]

**Proof.** Since \([Q'] \in \mathcal{I}_Q\), \( Q' \otimes_F \mathbb{V}_P \) is split by Proposition 3.2. Then, each class \([Q' \otimes_F K]\) obviously lies in \(\ker(h_2)\) from the commutative diagram:

\[
\begin{array}{c}
\text{Br}(F) \xrightarrow{\text{res}} \text{Br}(K) \\
\text{res} \downarrow \quad \quad \quad \text{res} \downarrow \\
\text{Br}(\mathbb{V}_P) \longrightarrow \text{Br}(\hat{K}_P)
\end{array}
\]

for each \( P \in \mathbb{P}(K/F) \). Here, the bottom map is the composition of the maps in (16). This proves one inclusion.

To verify the other inclusion, take any \([B] \in \ker(h_2)\). We claim that \([B]\) is a constant class. For each \( P \in \mathbb{P}(K/F) \), the class \([B]\) becomes trivial in \(\text{Br}(\hat{K}_P)\). By the definition of the ramification map in (15) associated to \( \mathbb{V} = \mathbb{V}_P \), we have \(\text{ram}([B]) = 0\). Recall (cf. [Ha, Lemma 2.11]) that the map \(\beta\) in (1) is the direct sum of these ramification maps where \( P \) ranges over all \( P \in \mathbb{P}(K/F) \). We then have \(\beta([B]) = 0\). It follows that \([B]\) is a constant class from the exact sequence (1). Let \([Q'] \in \text{Br}(F)\) such that \([Q' \otimes_F K] = [B]\). Since the bottom map in (18) is injective from (16), the class \([Q']\) becomes trivial in \(\text{Br}(\mathbb{V}_P)\). Replacing \(\text{Br}(\hat{K}_P)\) by \(\text{Br}(K\mathbb{V}_P)\) in (18), we still have a commutative diagram with all restriction maps. Consequently, \(B \otimes_K K\mathbb{V}_P\) is split for each \( P \), which implies that \([B] \in \ker(h)\). Hence, Theorem 3.7 shows that \([Q'] \in \mathcal{I}_Q\). \(\square\)

Putting Theorem 3.7, Theorem 4.4 and Theorem 4.5 all together, we have

\[
\ker(h) = \ker(h_1) = \ker(h_2).
\]
For $K = F(Q)$ as above, we say that the Hasse principle for $\text{Br}(K)$ holds (in the sense of $h_1$) if $\ker(h_1) = 0$. The following corollary is immediate from (19) since $|\ker(h)| = 2|\text{supp}(Q)|^{-2}$ by Theorem 3.7.

**Corollary 4.6.** Let $F$ be a global field. Suppose that $Q$ is a quaternion division algebra over $F$. Let $K = F(Q)$. Then, the Hasse principle for $\text{Br}(K)$ holds if and only if $|\text{supp}(Q)| = 2$.

**Example 4.7.** Let $p_1, p_2, \ldots, p_n$ be distinct odd prime numbers such that each $p_i \equiv 3 \pmod{4}$. According to Examples 3.3, 3.8, and 3.9, the Hasse principle holds for $\text{Br}(Q(x, \sqrt{-x^2 - 1}))$, for $\text{Br}(Q(x, \sqrt{-x^2 + p_1}))$, and for $\text{Br}(Q(x, \sqrt{-x^2 + p_1 p_2}))$. On the other hand, for $n \geq 3$, let $K = \text{Br}(Q(x, \sqrt{-x^2 + m}))$ where $m = p_1 p_2 \ldots p_n$. Then the nontrivial elements in

$$\ker(h_1) = \bigcap_{P \in \mathbb{F}(K/Q)} \text{Br}(K\mathbb{V}_P/K)$$

(see Examples 3.8 and 3.9) are the obstruction to the Hasse principle.

**Remark 4.8.** For $K = F(Q)$, assume that the Hasse principle for $\text{Br}(K)$ (in the sense of $h_1$) holds. However, there is no analogue to Hilbert’s Reciprocity Law in $K$ even for the constant classes. That is, there exist quaternion division algebras $B$ over $F(Q)$ such that $\widehat{\mathbb{F}}_p(Q)$ splits $Q$ for all $p \in P(F)$ but an odd number of prime spots. To see this, suppose $\text{supp}(Q) = \{p_1, p_2\}$. Take $q \in P(F)$ such that $q \neq p_i$ for $p = 1, 2$. By Hilbert’s Reciprocity Law 2.6, there exists a quaternion algebra $Q'$ over $F$ with $\text{supp}(Q') = \{q, p_i\}$. Put $B = Q' \otimes_F F(Q)$. Then, it is easy to check that $B \otimes_K \widehat{\mathbb{F}}_p(Q)$ is split for all $p \in P(F)$ except $q$.

**5. The cases when a finite number of prime spots are omitted.**

The purpose of this section is to generalize the results given in the previous sections. Specifically, we investigate the kernel of the map $h_1$ in (13) when a finite number of prime spots are dropped from $P(F)$. The need to consider what happens when finitely many primes are omitted arises often in algebraic number theory. In particular, when the author considered tractability of algebraic function fields of genus 0 over global fields, it was necessary to delete the dyadic prime spots (cf. Remark 5.10). In considering this relative $h_1$, we restrict our attention to the 2-torsion of the kernel of $h_1$ since otherwise $\ker(h_1)$ contains infinitely many elements in most cases as we will see in Remark 5.2. This part of $\ker(h_1)$ can be expressed in terms of quaternion algebras over $K$ possibly together with the images in $\text{Br}(K)$ of cyclic algebras of exponent 4 over $F$ (cf. Theorem 5.8).
Let us start this section with recalling the exact sequence (9) and the (local) results that

\[ \text{Br}(\hat{F}_p) \cong \begin{cases} \mathbb{Q}/\mathbb{Z} & \text{if } p \text{ is a finite prime spot,} \\ \mathbb{Z}/2\mathbb{Z} & \text{if } p \text{ is a real infinite,} \\ 0 & \text{if } p \text{ is a complex infinite.} \end{cases} \]

Thus, if \( F \) is an algebraic number field, we will consider only finite prime spots and real infinite spots on \( F \). (If \( F \) is an algebraic function field in one variable over a finite field, then every prime spot is finite.)

Now, we want to see what happens to \( \ker(i) \) in the exact sequence (9) if a finite number of prime spots in the direct sum are removed. For this, let \( S \) denote a finite subset of \( \mathcal{P}(F) \).

Define the map

\[ i_S : \text{Br}(F) \rightarrow \prod_{p \in P(F) - S} \text{Br}(\hat{F}_p). \] (20)

**Lemma 5.1.** For the map \( i_S \) in (20), we have the following:

(i) If \( |S| \leq 1 \), then \( i_S \) is injective.

(ii) If \( |S| \geq 2 \) and \( S \) contains only one finite prime spot (together with some real infinite spots), then we have

\[ \ker(i_S) = \{ [A] \mid A \text{ is a quaternion algebra over } F \text{ with supp}(A) \subseteq S \} \]

The cardinality of this set is \( 2^{n-1} \) where \( n = |S| \).

(iii) If \( S \) contains at least two finite prime spots, then \( \ker(i_S) \) is infinite.

**Proof.**

(i) If \( S = \emptyset \), then \( i_S = i \) in (9). If \( |S| = 1 \), \( i_S \) is still injective from the exactness at \( \bigoplus_p \text{Br}(\hat{F}_p) \) in (9).

(ii) Let \( A \) be a central division algebra over \( F \) with \( [A] \in \ker(i_S) \). The exactness at \( \bigoplus_p \text{Br}(\hat{F}_p) \) in (9) assures that \( [A] \) has a local invariant either 0 or \( \frac{1}{2} + \mathbb{Z} \) at each \( p \in S \) and further \( \text{supp}(A) \) has nonzero even cardinality. By Hilbert’s Reciprocity Law 2.6, \( A \) is a quaternion division algebra over \( \hat{F} \). This shows one inclusion and the other inclusion is clear. For the cardinality, apply Lemma 3.1.

(iii) It suffices to consider the case that

\[ S = \{ p_1, p_2 \mid p_1 \text{ and } p_2 \text{ are finite prime spots in } P(F) \}. \]

Utilizing the exactness at \( \bigoplus_p \text{Br}(\hat{F}_p) \) in (9) again, we observe that each \( n \)-torsion subgroup of \( \ker(i_S) \) has a class (of a cyclic algebra) of order \( n \) whose local invariant is \( \frac{k}{n} + \mathbb{Z} \) at \( p_1 \), \( \frac{n-k}{n} + \mathbb{Z} \) at \( p_2 \), and 0 otherwise, where \( n \geq 2 \) and \( 1 \leq k < n \).
**Remark 5.2.** Let $Q$ be a quaternion algebra over $F$ and let $K = F(Q)$. Assume that $S$ contains at least two finite prime spots as in Lemma 5.1 (iii). Define the map

$$h_S : Br(F(Q))' \longrightarrow \prod_{p \in P(F) - S} Br(\hat{F}_p(Q)).$$

Then for any $[A] \in \ker(i_S)$, each class $[A \otimes_F K]$ obviously lies in $\ker(h_S)$. This implies that $\ker(h_S)$ is infinite since $Br(K/F) \cong \{[F],[Q]\}$ by Proposition 2.1.

From now on, we focus on the 2-torsion subgroups of the Brauer groups above. For an abelian group $G$, let $2G$ denote the 2-torsion subgroup of $G$.

For a global field $F$, recall (cf. [Re, Theorem 32.19] and [Pi, Theorem, p. 236]) that any element in $2Br(F)$ is the class of a quaternion algebra over $F$.

Using (9) and the definition of the support, we obviously have the following exact sequence:

$$0 \longrightarrow \left\{ [Q'] \mid Q' \text{ is a quaternion algebra over } F \text{ with } \text{supp}(Q') \subseteq S \right\} \longrightarrow 2Br(F) \overset{i_S}{\longrightarrow} \prod_{p \in P(F) - S} 2Br(\hat{F}_p).$$

Let $F$ be any field. For any $a \in F$, there is an associated $(x - a)$-adic discrete valuation ring

$$F[x]_a := \{ f/g \mid f,g \in F[x], g(a) \neq 0 \},$$

whose residue field is $\overline{F[x]_a} = F$. Since there is a natural injection $Br(F[x]_a) \hookrightarrow Br(F(x))$, we can view $Br(F[x]_a)$ as a subgroup of $Br(F(x))$. The residue map $F[x]_a \to F$, called specialization at $a$, is given by $f(x)/g(x) \mapsto f(a)/g(a)$. This ring homomorphism induces the specialization map $Br(F[x]_a) \to Br(F)$, a group homomorphism. Recall that for any $[A] \in Br(F(x))$, we have $[A] \in Br(F[x]_a)$ for all but finitely many $a \in F$ (cf. [FSS, p. 924]).

**Lemma 5.3.** Let $F$ be a global field. Let $S$ be a finite subset of $P(F)$. If $[B] \in Br(F(x))'$ is not a constant class, then $[B]$ has a specialization $[A]$ in $Br(F)$ with $p \in \text{supp}(A)$ for some $p \in P(F) - S$.

**Proof.** There exist a finite degree extension field $E$ of $F$ and a prime spot $\mathfrak{p} \in P(E)$ such that for $p = \mathfrak{p}|_F \in P(F)$ we have $p \not\in S$ and that there is a specialization to $Br(F)$ such that the class $[B]$ specializes to some $[A]$ (which has the same order as $[B]$), and) whose support contains $p$ (see the proof of Theorem 2.4 and Theorem 2.5 in [FSS]).

\[\square\]
Proposition 5.4. Let $F$ be a global field and let $F(x)$ be the rational function field over $F$. Let $S$ be a finite subset of $P(F)$. Then, the following sequence is exact:

$$0 \longrightarrow \left\{ \left[ Q' \otimes F F(x) \right] \mid Q' \text{ is a quaternion algebra over } F \text{ with } \text{supp}(Q') \subseteq S \right\} \longrightarrow 2\text{Br}(F(x)) \overset{\tilde{j}_S}{\longrightarrow} \prod_{p \in P(F)-S} 2\text{Br}(\hat{F}_p(x)).$$

Proof. Let $[B] \in \ker(\tilde{j}_S)$ with $B$ nonsplit. We first show that $[B]$ is in fact a constant class. Otherwise, by Lemma 5.3, $[B] \in 2\text{Br}(F(x))$ has a specialization $[A] \in 2\text{Br}(F)$ and there exists $p \in P(F)$ with $p \in \text{supp}(A) - S$. Thus, for this $p$, $A \otimes_F \hat{F}_p$ is nonsplit and therefore $B \otimes_F \hat{F}_p(x)$ is also nonsplit from the commutative diagram:

$$
\begin{array}{ccc}
2\text{Br}(F(x)) & \longrightarrow & 2\text{Br}(\hat{F}_p(x)) \\
\downarrow^\rho & & \downarrow^\rho \\
2\text{Br}(F) & \longrightarrow & 2\text{Br}(\hat{F}_p)
\end{array}
$$

where the vertical maps $\rho$ are specialization maps. Hence, $[B] \notin \ker(\tilde{j}_S)$ if $[B]$ is not a constant class; so $[B]$ must be a constant class. Now, consider the following commutative diagram:

$$
\begin{array}{ccc}
2\text{Br}(F) & \longrightarrow & \prod_{p \in P(F)-S} 2\text{Br}(\hat{F}_p) \\
\downarrow & & \downarrow \\
2\text{Br}(F(x)) & \longrightarrow & \prod_{p \in P(F)-S} 2\text{Br}(\hat{F}_p(x)).
\end{array}
$$

Notice that the vertical maps are both injective. From the description of $\ker(\tilde{i}_S)$ in (21), it follows that $\ker(\tilde{j}_S) = \{ [Q' \otimes_F F(x)] \mid Q' \text{ is a quaternion algebra over } F \text{ with } \text{supp}(Q') \subseteq S \}$. $\square$

For a global field $F$, let $Q$ be a quaternion division algebra over $F$ with $\text{supp}(Q) = \{ p_1, p_2, \ldots, p_n \}$.

Let $K = F(Q)$. Consider another set $S = \{ q_1, q_2, \ldots, q_m \} \subseteq P(F)$ with $|S| = m$. We do not assume that $S$ is disjoint from $\text{supp}(Q)$. Define

$$(23) \quad \mathcal{F}_{Q,S} = \{ r \in F \mid r \notin \hat{F}_{p_1}^{\ast 2} \cup \cdots \cup \hat{F}_{p_n}^{\ast 2} \cup \hat{F}_{q_1}^{\ast 2} \cup \cdots \cup \hat{F}_{q_m}^{\ast 2} \}.$$
Recall from Proposition 2.7 that for each \( r \in \mathcal{F}_{Q,S} \), \( F(\sqrt{r}) \) is the residue field of a place of \( K/F \). We also define

\[
\mathcal{I}_{Q,S} = \left\{ [Q'] \mid Q' \text{ is a quaternion algebra over } F \text{ with } \supp(Q') \subseteq \supp(Q) \cup S \right\} \subseteq \text{Br}(F).
\]

The following lemma is a generalization of the last equality in Proposition 3.2.

**Lemma 5.5.** With notations as above, we have

\[
\bigcap_{r \in \mathcal{F}_{Q,S}} \text{Br}(F(\sqrt{r})/F) = \mathcal{I}_{Q,S}.
\]

The cardinality of this set is \( 2^{l-1} \) where \( l = |\supp(Q) \cup S| \).

**Lemma 5.6.** Let \( K = F(Q) \) as above. If \([B]\) is a nonconstant class in \( \text{Br}(K)' \), then \([B]\) is a nonconstant class in \( \text{Br}(F(\sqrt{r}))(Q) \) for some \( r \in \mathcal{F}_{Q,S} \).

**Proof.** We recall the commutative diagram (8) with exact rows. If \([B]\) is a nonconstant class in \( \text{Br}(K)' \), so in \( \text{Br}(F_{\text{sep}} - K/K) \), then \( \beta([B]) \) is nontrivial in \( \bigoplus_{P \in \mathcal{P}(K/F)} X(G_P) \). A little modification of Lemma 3.6 gives that if \( \chi_P \neq 0 \), then \( \chi_P|_{G_{P'}} \neq 0 \) for some \( r \in \mathcal{F}_{Q,S} \) and some \( P' \in \mathcal{P}(K(\sqrt{r})/F(\sqrt{r})) \) with \( P'|P \). For this \( r \in \mathcal{F}_{Q,S} \), \( (\oplus e_{P'} \cdot \text{res}) \circ \beta([B]) \neq 0 \) in \( \bigoplus_{P'|P} X(G_{P'}) \). Note that for each \( P \in \mathcal{P}(K/F) \), \( e_{P'} = 1 \) since \( v_P(r) = 0 \) as in the proof of Theorem 3.7. From the commutativity of the right square of (8), \( \text{res}([B]) \) is a nonconstant class in \( \text{Br}(K(\sqrt{r})). \)

For \( K = F(Q) \), we define a map

\[
\widetilde{h}_S : 2\text{Br}(K) \rightarrow \prod_{P \in \mathcal{P}(F) - S} 2\text{Br}(\widehat{F}_p(Q)).
\]

**Lemma 5.7.** Let \( K = F(Q) \) as above. Assume that \( S \) is a finite subset of \( P(F) \) and that \( \supp(Q) \not\subseteq S \). Let \( B \) be a nonsplit division algebra over \( K \). If \([B]\) is a constant class in \( \text{ker}(\widetilde{h}_S) \), then \( B \) is a quaternion algebra over \( K \). In fact, \([B] = [A \otimes_F K] \) where \( A \) is a quaternion algebra over \( F \).

**Proof.** Since \([B]\) \( \in \text{ker}(\widetilde{h}_S) \) is a constant class, we can choose a central division algebra \( A \) over \( F \) such that \([A \otimes_F K] = [B]\). Since \( \text{exp}(B) = 2 \), the function field \( K \) splits \( A^{\otimes 2} := A \otimes_F A \). We claim that \([A^{\otimes 2}]\) is trivial. Suppose that \([A^{\otimes 2}]\) were nontrivial. Note then that \( A^{\otimes 2} \sim Q \) by Proposition 2.1. Take \( p \in \supp(Q) - S \). For this \( p \), we have the following
commutative diagram:

\[
\begin{array}{ccc}
\text{Br}(F) & \xrightarrow{\text{res}} & \text{Br}(\hat{F}_p) \\
\text{res} & & \text{res} \\
\downarrow & & \downarrow \\
\text{Br}(F(Q)) & \xrightarrow{\text{res}} & \text{Br}(\hat{F}_p(Q))
\end{array}
\]

(26)

Observe that \(\exp(A \otimes F \hat{F}_p) = 4\) since \(Q \otimes F \hat{F}_p \sim A^{o2} \otimes F \hat{F}_p\) has exponent 2. It follows that \(A \otimes F \hat{F}_p \otimes F \hat{F}_p\) is nonsplit over \(\hat{F}_p(Q)\) by Proposition 2.1. However, the diagram (26) shows that this fact contradicts our assumption that \([B] \in \ker(\hat{h}_S)\). Hence, \([A^{o2}]\) must be trivial. This shows that \(\exp(A) \leq 2\). Since \(F\) is a global field, we have \(\text{ind}(A) = \exp(A)\) and so \(A\) is a quaternion algebra over \(F\) (cf. [Pi, Theorem, p. 236]). Therefore \(B\) is a quaternion algebra over \(K\). □

We need to generalize the notion of the support to central simple algebras, not restricting to quaternion algebras, for our main theorem in this section. For a central simple algebra \(A\) over a global field \(F\), define

\[
\text{supp}(A) = \{ p \in P(F) \mid A \otimes F \hat{F}_p \text{ is nonsplit} \}.
\]

**Theorem 5.8.** Let \(F\) be a global field and \(K = F(Q)\) where \(Q\) is a quaternion division algebra over \(F\). Assume that \(S\) is a finite subset of \(P(F)\). For the \(\hat{h}_S\) in (25), we have the following:

(i) If \(\text{supp}(Q) \not\subseteq S\), then

\[
\ker(\hat{h}_S) = \bigcap_{r \in F_{Q,S}} \text{Br}(K(\sqrt{r})/K) = \{ [Q' \otimes_F K] \mid Q' \in I_{Q,S} \}
\]

for the \(F_{Q,S}\) in (23), and the \(I_{Q,S}\) in (24). The cardinality of \(\ker(\hat{h}_S)\) is \(2^{n-2}\) where \(n = |\text{supp}(Q) \cup S|\).

(ii) If \(\text{supp}(Q) \subseteq S\), then

\[
\ker(\hat{h}_S) = \left\{ [Q' \otimes_F K] \mid Q' \text{ is a quaternion algebra over } F \text{ with } \text{supp}(Q') \subseteq S \right\} \cup \left\{ [A \otimes_F K] \mid A^{o2} \sim Q \text{ and } \text{supp}(A) \subseteq S \right\}.
\]

The cardinality of \(\ker(\hat{h}_S)\) is \(2^{n-1}\) where \(n = |S|\).

**Proof.** (i) Using a similar argument to that in the proof of Theorem 3.7, we have the second equality in (27). We now want to show the first equality. Let

\[
\mathcal{I} = \{ [Q' \otimes_F K] \mid Q' \in I_{Q,S} \}.
\]

Plainly, \(\mathcal{I} \subseteq \ker(\hat{h}_S)\). For the other inclusion, assume that \([B] \in \ker(\hat{h}_S)\).
We first assert that \([B]\) is a constant class. To see this, suppose that \([B]\) were a nonconstant class. Then, \([B] \notin \mathcal{I} = \bigcap_{r \in \mathcal{F}_{Q,S}} \text{Br}(K(\sqrt{r})/K)\). By Lemma 5.6, we can find \(r \in \mathcal{F}_{Q,S}\) so that \([B \otimes_K K(\sqrt{r})]\) is still a nonconstant class in \(\text{Br}(K(\sqrt{r}))\). Put \(E = F(\sqrt{r})\). Let \(T\) be the set of the extensions of the prime spots in \(S\) to the quadratic extension \(E\). Consider the following diagram:

\[
\begin{array}{ccc}
2\text{Br}(F(Q)) & \xrightarrow{\sim_{S}} & \prod_{p \in \mathcal{P}(E) - T} 2\text{Br}(\hat{F}_p(Q)) \\
\downarrow & & \downarrow \\
2\text{Br}(E(Q)) & \xrightarrow{\sim_T} & \prod_{p \in \mathcal{P}(E) - T} 2\text{Br}(\hat{E}_p(Q))
\end{array}
\]

(29)

Note that this diagram is commutative (cf. (14)). Since \(E\) splits \(Q\), the function fields \(E(Q)\) and \(\hat{E}_p(Q)\) are both rational over \(E\) and \(\hat{E}_p\) respectively. Then,

\[
\ker(\tilde{h}_T) = \{[Q' \otimes_K E(Q)] \mid \text{supp}(Q') \subseteq T\}
\]

by the exact sequence (22) (replacing \(F\) by \(E\)). Since \(\ker(\tilde{h}_T)\) contains only constant classes, it follows that \([B \otimes_K E(Q)] \notin \ker(\tilde{h}_T)\) and so \([B] \notin \ker(\tilde{h}_S)\) from the commutativity of (29). This contradicts our assumption. Therefore \([B]\) is a constant class.

Since \([B]\) is a constant class, Lemma 5.7 shows that \([B] = [Q' \otimes_F K]\) for some quaternion algebra \(Q'\) over \(F\). We now claim that \([B] \notin \mathcal{I}\). Otherwise, we can choose a prime spot \(p \in \text{supp}(Q') - (\text{supp}(Q) \cup S)\). For this \(p\), \(Q' \otimes \hat{F}_p\) is nonsplit and so is \(Q' \otimes \hat{F}_p(Q)\) since \(Q\) is split over \(\hat{F}_p\) (cf. Lemma 2.2). Hence, \([B] \notin \ker(\tilde{h}_S)\), which is a contradiction. This proves the first equality in (27).

From the definition of \(\mathcal{I}_{Q,S}\) in (24) together with Proposition 2.1, it is clear that \(|\ker(\tilde{h}_S)| = 2^{n-2}\) where \(n = |\text{supp}(Q) \cup S|\).

(ii) One inclusion \((\supseteq)\) in (28) is clear. We show the other inclusion \((\subseteq)\). By the same argument as in the proof of (i), any \([B] \in \ker(\tilde{h}_S)\) is a constant class so that we can find \([A] \in \text{Br}(F)\) with \([B] = [A \otimes_F K]\). Consider the following commutative diagram:

\[
\begin{array}{ccc}
\text{Br}(F) & \xrightarrow{i_S} & \prod_{p \in \mathcal{P}(F) - S} \text{Br}(\hat{F}_p) \\
\downarrow & & \downarrow \\
\text{Br}(F(Q)) & \xrightarrow{h_S} & \prod_{p \in \mathcal{P}(F) - S} \text{Br}(\hat{F}_p(Q)).
\end{array}
\]

(30)
Notice that for each \( p \notin S \), \( \widehat{F}_p(Q) \) is purely transcendental over \( \widehat{F}_p \) and so the right vertical map in (30) is injective. This implies that \( [A] \in \ker(i_S) \). Since \( \exp(B) = 2 \), we have \( \exp(A) = 2 \) or 4 because \( K \) is a quadratic extension of a purely transcendental extension of \( F \). This implies that \( \ind(A) = 2 \) or 4 since \( F \) is a global field. If \( \ind(A) = 2 \), then \( A \) is a quaternion algebra with \( [A] \in \ker(i_S) \), so \( \supp(A) \subseteq S \) by (21). If \( \ind(A) = 4 \), then \( A^{\otimes 2} \) is nonsplit over \( F \), but split over \( F(Q) \). Therefore, \( A^{\otimes 2} \sim Q \) by Proposition 2.1. The injectivity of the right vertical map in (30) shows that \( \supp(A) \subseteq S \).

Finally, for the cardinality of \( \ker(h_S) \), let \( H_1 \) denote the set of preimages in \( \Br(F) \) of the first set of the union in (28) and \( H_2 \) denote that of the second set. Notice that \( H_1 \) is a subgroup of \( \Br(F) \). We show that \( H_2 \) is a coset relative to \( H_1 \) generated by any element \( [A] \) in \( H_2 \). In fact, for any two \( [A], [A'] \in H_2 \) with \( A \neq A' \), we have

\[
(A' \otimes_F A'^{\text{op}})^{\otimes 2} \sim Q \otimes F Q'^{\text{op}} \sim F,
\]

where \( A'^{\text{op}} \) is the opposite algebra of \( A \). It follows that \( A' \otimes_F A'^{\text{op}} \) has exponent 2. Evidently, \( \supp(A' \otimes_F A'^{\text{op}}) \subseteq S \). Thus, \( A' \otimes_F A'^{\text{op}} \sim \tilde{Q} \) for some \( [\tilde{Q}] \in H_1 \). That is, \( A' \sim A \otimes_F \tilde{Q} \). On the other hand, it is obvious that for any \( [Q'] \in H_1, [A \otimes_F Q'] \in H_2 \). Hence, \( |H_1| = |H_2| = 2^{n-1} \) where \( n = |S| \). Consequently,

\[
|\ker(h_S)| = \frac{1}{2} (2^{n-1} + 2^{n-1}) = 2^{n-1}.
\]

This completes the proof.

\[\square\]

**Remark 5.9.** Suppose that \( \supp(Q) \subseteq S \). If the cardinality of the set \( \supp(Q) \cup S \) is even, there exists a quaternion algebra \( Q' \) such that \( \supp(Q') = \supp(Q) \cup S \) by Hilbert’s Reciprocity Law 2.6. Then,

\[
\ker(h_S) = \ker(h_1),
\]

where \( h_1 \) is a map in (13) for \( K = F(Q') \).

**Remark 5.10.** Suppose that \( Q \) is a quaternion division algebra with \( \supp(Q) = \{ p_1, p_2 \} \) and \( S = \{ q \} \). For the map

\[
(31) \quad h_S : \Br(F(Q))' \longrightarrow \prod_{p \in P(F)-S} \Br(\widehat{F}_p(Q)),
\]

Proposition 2.1 and Lemma 5.1 (i) allow us to conclude that

\[
\ker(h_S) = \ker(h_S) = \left\{ 0, [Q' \otimes_F K] \mid \text{Q’ is a quaternion algebra over } F \text{ with } \supp(Q') = \{ p_1, q \} \right\}.
\]

Observe that if \( q \) is equal to one of \( p_i \), then \( \ker(h_S) \) is trivial. This situation arose in [Ha] when we studied the tractability of function fields of genus 0 over global fields as follows: Let \( F \) be an algebraic number field with exactly one dyadic spot \( \mathfrak{d} \). Assume that \( \supp(Q) = \{ p_1, \mathfrak{d} \} \) and \( S = \{ \mathfrak{d} \} \). Then,
the map $h_S$ is injective. On the other hand, let $F$ be an algebraic function field in one variable over a finite field with $\mathcal{S} = \varnothing$. Then

$$\ker(h_S) = \ker(h_1) = 0,$$

where $h_1$ is the map in (13), since $\text{supp}(Q)$ contains exactly 2 prime spots. In both cases, the injectivity of $h_S$ guarantees that the function field $F(Q)$ is tractable since $\widehat{F}_p(Q)$ is tractable for each nondyadic local field $\widehat{F}_p$ (see [Ha, Theorem 4.9] for details).

**Example 5.11.** Let $Q = (-1, p_1p_2/Q)$, where $p_1$ and $p_2$ are distinct odd primes with $p_i \equiv 3 \pmod{4}$. Let $K = \mathbb{Q}(Q) = \mathbb{Q}(x, \sqrt{-x^2 + p_1p_2})$. Then $\text{supp}(Q) = \{p_1, p_2\}$ as in Example 3.5. Assume that $S = \{2, \infty\}$ where 2 is the dyadic spot and $\infty$ is the real infinite spot of $\mathbb{Q}$. Then,

$$\ker(\widetilde{h}_S) = \{0, \left[-1, (-1/Q)\right], \left[(-1, p_1/Q)\right], \left[(-1, p_2/Q)\right]\}.$$

We can also describe the kernel of the map $h_S$ in (31). In fact, since $\infty$ is the real infinite spot, any Brauer class in $\ker(h_S)$ has local invariant either 0 or $\frac{1}{2} + \mathbb{Z}$ at the dyadic spot 2 (see Lemma 5.1 (ii)). It follows that

$$\ker(h_S) = \ker(\widetilde{h}_S).$$

Moreover, by Remark 5.9 we have $\ker(\widetilde{h}_S) = \ker(h_1)$ where $h_1$ is the map in (13) for $K = \mathbb{Q}(x, \sqrt{-x^2 - p_1p_2})$.

In case that $\text{supp}(Q) \subseteq \mathcal{S}$, the following remark provides us with concrete descriptions of the inverse images in $\text{Br}(F)$ of the elements in $\ker(\widetilde{h}_S)$.

**Remark 5.12.** For $[A \otimes_F K] \in \ker(\widetilde{h}_S)$ in Theorem 5.8 (ii), the local invariant of $A$ at $p \in P(F)$ is as follows:

$$\text{inv}_p(A) = \begin{cases} \frac{1}{4} \text{ or } -\frac{1}{4} + \mathbb{Z} & \text{if } p \in \text{supp}(Q) \\ 0 \text{ or } \frac{1}{2} + \mathbb{Z} & \text{if } p \in \mathcal{S} - \text{supp}(Q) \\ 0 & \text{if } p \not\in \mathcal{S}. \end{cases}$$

Note that $A \otimes_F K$ has exponent 2 but index 4. In fact, by the index reduction formula since $K$ is a generic splitting field of $Q$ (cf. [SV, Theorem 2.3, p. 735]), we obtain

$$\text{ind}(A \otimes_F K) = \min(\text{ind}(A), \text{ind}(A \otimes_F Q)) = \min(\exp(A), \exp(A \otimes_F Q)) = 4.$$ 

This tells us that $A \otimes_F K$ is a product of two quaternion algebras over $K$ by a theorem of Albert (cf. [KMRT, Theorem 16.1, p. 233]). Moreover, we observe that neither of these quaternion algebras is of constant class. For, if one of them were of constant class then so would be the other. This would imply that $A$ is a tensor product of quaternion algebras over $\mathbb{F}$ contradicting the fact that $\exp(A) = 4$. 


Example 5.13. Let $K = F(Q)$ as above. Suppose that $\text{supp}(Q) = \{p_1, p_2\}$ and $S = \{p_1, p_2, p_3, p_4\}$. Then $|\ker(\hat{h}_S)| = 8$. To see this, let $A := (n_1, n_2, n_3, n_4)$ denote a cyclic algebra over $F$ with local invariants

$$\text{inv}_P(A) = \begin{cases} n_i + \mathbb{Z} & \text{if } p = p_i \ (i = 1, \ldots, 4), \\ 0 & \text{if } p \not\in S. \end{cases}$$

Obviously, we have $\text{supp}(A) \subseteq S$. The preimages in $\text{Br}(F)$ of $\ker(\hat{h}_S)$ are the classes of the following algebras:

- $Q_1 = (0, 0, 0, 0)$
- $Q_2 = (0, \frac{1}{2}, \frac{1}{2}, 0)$
- $Q_3 = (0, \frac{1}{2}, 0, \frac{1}{2})$
- $Q_4 = (0, 0, \frac{1}{2}, \frac{1}{2})$
- $A_1 = (\frac{1}{4}, \frac{1}{4}, 0, 0)$
- $A_2 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0)$
- $A_3 = (\frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4})$
- $A_4 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$

Here, $Q_i$ are quaternion algebras over $F$, and $A_i$ are cyclic algebras over $F$ of exponent 4 satisfying $A_i \otimes_F Q_i + 4 \sim Q_i$. Notice that

$$Q_i + 4 \sim Q_i \otimes_F Q \text{ and } A_i + 4 \sim A_i \otimes_F Q \text{ for } i = 1, \ldots, 4.$$  

By Theorem 5.8, we have

$$\ker(\hat{h}_S) = \{[Q_i \otimes_F K], [A_i \otimes_F K] \mid i = 1, \ldots, 4\}.$$  

References


<table>
<thead>
<tr>
<th>Reference</th>
<th>Authors</th>
<th>Title</th>
<th>Source</th>
<th>Volume</th>
<th>Pages</th>
<th>MR Number</th>
<th>Zbl Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>[FS]</td>
<td>B. Fein and M. Schacher</td>
<td>Brauer groups of rational function fields over global fields</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Ha]</td>
<td>I. Han</td>
<td>Tractability of algebraic function fields in one variable of genus zero over global fields</td>
<td>J. Algebra</td>
<td>244</td>
<td>(2001), 217-235</td>
<td>MR 2002k:12003</td>
<td></td>
</tr>
<tr>
<td>[OM]</td>
<td>O.T. O’Meara</td>
<td>Introduction to Quadratic Forms</td>
<td>Springer, Berlin</td>
<td>1963</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Received July 27, 2001 and revised November 27, 2001.

California State University
San Bernardino CA 92407-2397
E-mail address: ihan@math.csusb.edu
KNOTTED CONTRACTIBLE 4-MANIFOLDS IN $S^4$

W.B.R. Lickorish

Examples are given to show that some compact contractible 4-manifolds can be knotted in the 4-sphere. It is then proved that any finitely presented perfect group with a balanced presentation is a knot group for an embedding of some contractible 4-manifold in $S^4$.

1. Introduction.

A construction will here be described that can produce a compact contractible 4-manifold $M$ embedded piecewise linearly (or smoothly) in $S^4$ with the fundamental group of its complement being nontrivial. Then, another embedding of $M$ in $S^4$ will be produced which has simply connected complement. Several examples of this will be given. Of course, the construction emphasizes that contractible spaces do not behave entirely as do single points. It is important to note that these embeddings are piecewise linear or smooth; they are certainly not wild. The famous construction of the Alexander wild horned sphere gives a wild embedding of a 3-ball in $S^3$ that has its complement not simply connected. However the boundary of a contractible compact 3-manifold is just a 2-sphere, so by the piecewise linear 3-dimensional Schönflies theorem, if such a manifold can be embedded piecewise linearly in $S^3$, each of the manifold and its complement must be a 3-ball.

Recall the general definition of knotting, when all maps and spaces are in the piecewise linear category: A polyhedron $X$ knots in a polyhedron $Y$ if there are two embeddings, $e_0$ and $e_1$ of $X$ in $Y$, that are homotopic but not ambient isotopic. The embeddings are ambient isotopic if there exist homeomorphisms $F_t : Y \rightarrow Y$, for each $t \in [0, 1]$, such that $(y, t) \mapsto (F_t(y), t)$ defines a piecewise linear homeomorphism from $Y \times [0, 1]$ to itself, $F_0$ is the identity and $F_1 e_0 = e_1$. Thus to be ambient isotopic the complements of the images of the two embeddings must certainly be homeomorphic. The knotting phenomenon explores the possibility of moving between embeddings along a path of embeddings as opposed to moving along a path of maps. The examples given here are of contractible 4-manifolds that can knot in $S^4$ for, just as in classical knot theory, the fundamental group of complements is used to show embeddings are not ambient isotopic. Examples of knots usually rely on the entwining of some nontrivial cycle, but here there
is none. In fact, in higher dimensions, if $X$ and $Y$ are piecewise linear manifolds with $\dim Y - \dim X \geq 3$, there are theorems of Hudson [4] that assert that there is no knotting of $X$ in $Y$ provided these spaces are sufficiently highly connected.

It should be noted that when $M$ is a contractible 4-manifold, piecewise linearly contained in $S^4$, the Alexander duality theorem implies that $S^4 - M$ has the same homology as a point. Thus $\pi_1(S^4 - M)$ is a perfect group in contrast to the situation of classical knot theory. It will be proved in Theorem 3 that for any finitely presented perfect group with a balanced presentation (that is, a presentation with the same number of generators as relators) there are embeddings $e_0$ and $e_1$ of some contractible 4-manifold $M$ into $S^4$ so that $\pi_1(S^4 - e_1 M)$ is the given group and $\pi_1(S^4 - e_0 M)$ is trivial.

The author is grateful to Simon Norton for making a helpful remark and to Charles Livingston for a correction. Livingston has extended the method of this paper to obtain, for certain contractible 4-manifolds, infinitely many knots in $S^4$.

2. Examples of the embedding construction.

The theorem that now follows is really an example describing the main simple idea of the construction of this paper. The second theorem amplifies it to more general circumstances.

**Theorem 1.** There are two piecewise linear (or smooth) embeddings, $e_0$ and $e_1$, of a certain compact contractible 4-manifold $M$ into $S^4$ such that $\pi_1(S^4 - e_1 M)$ is nontrivial and $S^4 - e_0 M$ is contractible.

**Proof.** Firstly, construct a compact 4-manifold $X$ by adding three 1-handles and three 2-handles onto a 4-ball in the following way. The handles are to be chosen so that $\pi_1(X)$ has the presentation

$$\langle a, b, c : b^{-1}c^{-2}bc^3, c^{-1}a^{-2}ca^3, a^{-1}b^{-2}ab^3 \rangle,$$

where based loops encircling the three 1-handles represent $a$, $b$ and $c$, and the attaching circles of the three 2-handles give the three relators. This situation is shown in Figure 1 in the notation common in considerations of the ‘Kirby calculus’ (see [2] for example).

The diagram of Figure 1 shows curves in the 3-sphere, the boundary of the 4-ball. Open regular neighbourhoods, of three standard disjoint discs in the 4-ball, are to be removed from the 4-ball to create a ball with the three 1-handles added. The boundaries of these discs are the circles, decorated with dots, labeled $a$, $b$ and $c$. That this is in order can be checked as follows. A ball with 1-handles added can be changed back to a ball by adding 2-handles to cancel the 1-handles; removing those 2-handles consists of removing neighbourhoods of the discs that are the co-cores of the 2-handles. Thus a 4-ball, with 1-handles added, is the same as a 4-ball from
which standard 2-handles have been removed. A 1-handle can be regarded as $D^1 \times D^3$ with $\partial D^1 \times \ast$ being the attaching sphere and $\ast \times \partial D^3$ being the belt sphere (where each $\ast$ is a base point). In Figure 1 a belt sphere consists of the union of a disc spanning a dotted circle, less a regular neighbourhood of that circle, and a disc in the boundary of the 2-handle that has been removed. Meridians encircling the three dotted circles represent generators, to be called $a$, $b$ and $c$, of the fundamental group of the ball with 1-handles, and a based closed curve represents a word, in $a$, $b$ and $c$, corresponding to its signed intersections with the three belt spheres. In this way the curves shown, labeled $\alpha$, $\beta$ and $\gamma$, represent $b^{-1}c^{-2}bc^3$, $c^{-1}a^{-2}ca^3$ and $a^{-1}b^{-2}ab^3$. Thus adding 2-handles, with these curves as attaching spheres (choose the zero framings), gives the 4-manifold $X$ with the required presentation for $\pi_1(X)$. It has been shown by Rapaport [6] that this is the presentation of a nontrivial group. There are, of course, very many ways that attaching curves can be chosen for the 2-handles in order to achieve this presentation (and the choice will be explored further in Theorem 3), but the one shown is about the simplest and is the one that will now be considered. The situation is shown schematically in Figure 2.
The manifold \( X \) is a 4-ball, \( B_1 \) say, with 2-handles removed and 2-handles added. Regard this 4-ball as being contained in \( S^4 \) and consider the complementary 4-ball \( B_2 \). The 2-handles removed from \( B_1 \) can be thought of as added to \( B_2 \). The other 2-handles, that were added to \( B_1 \), were added with zero framing along unknotted, unlinked curves (labeled \( \alpha \), \( \beta \) and \( \gamma \)), so they can be regarded as standard 2-handles removed from \( B_2 \). Thus the closure of \( S^4 - X \) is the 4-ball \( B_2 \) with three 2-handles removed (creating added 1-handles) and three 2-handles added and \( \text{this} \) is to be the required 4-manifold \( M \). The situation for \( M \) is again represented by Figure 1, except that now the dots should be removed from the curves labeled \( a \), \( b \) and \( c \) and placed on those labeled \( \alpha \), \( \beta \) and \( \gamma \). However, the \( \alpha \)-curve bounds a disc that meets only the \( c \)-curve; there are similar discs for the \( \beta \)- and \( \gamma \)-curves. The words in \( \alpha \), \( \beta \) and \( \gamma \), coming from the intersections of the \( a \), \( b \) and \( c \) curves with these discs, make it clear that \( \pi_1(M) \) is presented by

\[
\langle \alpha, \beta, \gamma : \alpha^3 \alpha^{-2}, \beta^3 \beta^{-2}, \gamma^3 \gamma^{-2} \rangle
\]

which, very obviously, presents the trivial group. Thus \( M \) is simply connected. A count of the handles shows that the Euler characteristic of \( M \) is 1, hence \( H_2(M) = 0 \). Furthermore \( H_r(M) = 0 \) for \( r > 2 \), as there are no \( r \)-handles for \( r > 2 \), and so, by the Hurewicz isomorphism theorem, \( M \) has all homotopy groups trivial and hence is contractible. Note that \( \pi_1(\partial M) \neq \{1\} \), as otherwise \( \pi_1(X) \cong \pi_1(S^4) \) by the Van Kampen theorem. Hence \( M \) is not a 4-ball. The inclusion of \( M \), as so defined in \( S^4 \), is the embedding \( e_1 \).

The above presentation given for \( \pi_1(M) \) coming from the handle structure of \( M \) is almost trivial. It certainly reduces to the trivial presentation by Andrews-Curtis moves (see below). Any such \( M \) has the property that \( M \times [0, 1] \cong B^5 \) where \( B^5 \) is a 5-ball. To show that in this instance, it is necessary only to realise that \( M \times [0, 1] \) has the same handle structure as does \( M \). The extra dimension means that, when a 2-handle is attached (to the boundary of a 5-manifold) only the homotopy class of the attaching map is significant (a homotopy of attaching circles can be changed to an isotopy by using the fourth dimension to prevent the circles from crossing each other).

Now let \( e_0 \) be the inclusion of \( M \times \{0\} \) in the 4-sphere \( \partial(M \times [0, 1]) \). The complement of \( M \times \{0\} \) in this sphere is \( (\partial M \times [0, 1]) \cup M \times \{1\} \) and this is just another copy of the contractible manifold \( M \).

For a second example consider \( \langle a, b : ab^2ab^{-1}, a^4ba^{-1}b \rangle \), a presentation of the perfect group \( G \) of 120 elements that is the fundamental group of the Poincaré homology 3-sphere. The method of the proof of Theorem 1 constructs a 4-manifold \( X \subset S^4 \) with \( \pi_1(X) \cong G \) and with the fundamental group of the corresponding \( M \) presented by \( \langle \alpha, \beta : \alpha^2\beta^3, \alpha^{-1}\beta^{-2} \rangle \). Again this easily reduces to the trivial presentation by Andrews-Curtis moves so that \( M \times I \) is a 5-ball.
The above construction works easily for the presentations
\[ \langle a_1, a_2, \ldots, a_m : r_1, r_2, \ldots, r_m \rangle \]
for every \( m \geq 4 \) when \( r_i = a_i^{-1}a_{i+1}a_i^2a_{i+1}^{-1} \) for \( i = 1, 2, \ldots m \) modulo \( m \), and also when \( r_i = a_i^{-1}a_{i+1}^{-2}a_i^3a_{i+1} \). These are known to be presentations of infinite groups (see [3] and [5]).

Note that \( \langle a, b : a^{-1}b^{-2}ab^3, b^{-1}a^{-2}ba^3 \rangle \) is a presentation of the trivial group. If \( M \) is constructed from this presentation for \( X \) it is not clear whether the embedding of \( M \) is in any sense knotted.


The above proof makes a brief mention of the Andrews-Curtis moves. These moves are elementary changes that can be made to a group presentation that do not alter the group that is presented. The moves are also called ‘extended Nielsen transformations’ in [1], they are called ‘\( Q \)-transformations’ in [6] and they are sometimes also called ‘Markov operations’. The permitted changes to a presentation \( \langle a_1, a_2, \ldots, a_m : r_1, r_2, \ldots, r_n \rangle \) are the following moves and the inverses of these moves:

(i) Change \( r_i \) to \( r_ia_ja_j^{-1} \) or \( r_i a_j^{-1}a_j \).

(ii) Change \( r_i \) to a cyclic permutation of \( r_i \).

(iii) Change \( r_i \) to \( r_i^{-1} \).

(iv) Change \( r_i \) to \( r_i r_j \) where \( j \neq i \).

(v) Add a new generator \( a_{m+1} \) and a new relator \( a_{m+1}w \) where \( w \) is a word in \( a_1, a_2, \ldots, a_m \).

These are precisely the moves that can easily be imitated on a 5-manifold comprised of 0-handles, 1-handles and 2-handles only. If the handles of such a 5-manifold correspond to a presentation of the trivial group that can be reduced to the trivial presentation (that is, the empty presentation) by the above moves and their inverses, then it is shown in [1] that the manifold is the 5-ball. It is this result that is used in the above proof. The Andrews-Curtis conjecture [1] is that any presentation of the trivial group be reducible to the trivial presentation by the above moves and their inverses. This is popularly thought to be false, \( \langle a, b : a^{-1}b^{-2}ab^3, b^{-1}a^{-2}ba^3 \rangle \) being one of many proposed counter-examples. The truth of the Andrews-Curtis conjecture would imply the truth of another conjecture that asserts that any 5-dimensional regular neighbourhood of a contractible 2-complex be a 5-ball (such a neighbourhood is known to be unique).

4. Arbitrary finitely presented perfect groups.

A few simple remarks lead up to an elementary, but possibly surprising, little lemma about finitely presented perfect groups and presentations of
the trivial group. Suppose that an abelian group $E$ is freely generated as an abelian group (with additive notation) by the generators $e_1, e_2, \ldots, e_m$. The quotient of $E$ by the subgroup generated by the $n$ elements $\{\sum_{j=1}^m a_{ij} e_j : i = 1, 2, \ldots, n\}$ is said to be presented by the $n \times m$ integer matrix $A = \{a_{ij}\}$. When $m = n$ the quotient is the trivial group if and only if $A$ is unimodular, that is, $\det A = \pm 1$. A presentation $P$ of any group (in multiplicative notation) leads at once to a presentation of the abelianisation of that group, by just deciding that all symbols commute. It is then sensible in each relator to assemble together all occurrences of a generator and its inverse, cancelling where possible, to obtain from the resulting exponents in each relator a presentation matrix $A$ of the abelianisation of the group. The following lemma considers such things in the reverse order, showing that, if $A$ presents the trivial abelian group, then $P$ can be chosen to present the trivial group.

**Lemma 2.** Suppose that $A$ is a unimodular $n \times n$ matrix of integers. Then there exists a presentation $P$ of the trivial group that has $A$ as its abelianised presentation matrix. Furthermore, $P$ is equivalent to the trivial presentation by Andrews-Curtis moves.

**Proof.** Starting from the identity $n \times n$ matrix, the unimodular matrix $A$ can be created by a sequence of row operations in which either a row is multiplied by $-1$ or a row is added to another row. These moves can be mimicked by changes to a presentation $\langle a_1, a_2, \ldots, a_n : r_1, r_2, \ldots, r_n \rangle$ of the trivial group, where initially $r_i = a_i$ for each $i$. If the $i$th row of the matrix is multiplied by $-1$, change $r_i$ to $r_i^{-1}$; if row $i$ is added to row $j$ then change $r_j$ to $r_j r_i$. At each stage the presentation is of the trivial group and at each stage the matrix is the corresponding presentation matrix of the abelianised group. Of course the moves used on the presentation are all Andrews-Curtis moves.

**Theorem 3.** Let $G$ be any finitely presented perfect group having a balanced presentation. Then there is a compact contractible 4-manifold $M$ contained in $S^4$ such that $\pi_1(S^4 - M) \cong G$ and $M \times I$ is a 5-ball (so that, if $G$ is non-trivial, there is a distinct second embedding of $M$ in $S^4$ having contractible complement).

**Proof.** Let $\langle a_1, a_2, \ldots, a_n : r_1, r_2, \ldots, r_n \rangle$ be a presentation $P$ of $G$. The construction proceeds as in the proof of Theorem 1. Remove from the 4-ball $B^4$ neighbourhoods of $n$ standard disjoint spanning discs to create a ball with $n$ 1-handles added. The boundaries of the discs form a set of $n$ unlinked simple closed (‘dotted’) curves in $\partial B^4 = S^3$, which are labeled $a_1, a_2, \ldots, a_n$. In the following way construct, as the boundaries of disjoint discs $D_1, D_2, \ldots, D_n$ contained in $S^3$, simple closed curves $\alpha_1, \alpha_2, \ldots, \alpha_n$, corresponding to $r_1, r_2, \ldots, r_n$, which are to be the attaching circles for $n$ 2-handles. Begin with small, disjoint, oriented discs $\Delta_1, \Delta_2, \ldots, \Delta_n$ in the
complement of \( a_1 \cup a_2 \cup \cdots \cup a_n \). For each letter \( a_i \pm 1 \) in the word \( r_1 \) take a small meridian disc of the curve \( a_i \), oriented according to the exponent on
the letter, and to construct \( D_1 \), join the boundaries of these meridian discs by thin bands to the boundary of \( \Delta_1 \), in the order around \( \partial \Delta_1 \) specified by
\( r_1 \). The discs \( D_2, \ldots, D_n \) are constructed similarly from \( r_2, \ldots, r_n \) and there
is no difficulty in ensuring that the \( D_i \) are embedded and mutually disjoint.

As in Theorem 1, form a 4-manifold \( X \subset S^4 \) by adding \( n \) 2-handles
with zero framing along \( \alpha_1, \alpha_2, \ldots, \alpha_n \) to the ball with \( n \) 1-handles. Then
\( \pi_1(X) \cong G \). The key point to note now is that the meridian discs described
above (for all the \( D_i \) together) can be taken in any order around \( a_i \). Different
choices of order probably give different manifolds \( X \) and \( M \), where again
\( M \) is the closure of \( S^3 - X \). Also note that the given presentation of \( G \) can
be amended by the insertion of any number of copies of \( a_i a_i^{-1} \), for any \( i \), into
any \( r_j \) without changing \( G \) nor the presentation matrix \( A \) of the (trivial)
abelianisation of \( G \) coming from that presentation. Now \( \pi_1(M) \) has a
presentation \( \Pi \) of the form \( \langle \alpha_1, \alpha_2, \ldots, \alpha_n : \rho_1, \rho_2, \ldots, \rho_n \rangle \) where the relators
record in order the occurrence of the meridian discs around \( a_1, a_2, \ldots, a_n \)
(each signed intersection of \( a_i \) with a meridian disc contained in \( D_j \) produc-
ing an \( a_i^{\pm 1} \) entry in \( \rho_j \)). The abelian presentation matrix coming from \( \Pi \) is
the transpose of \( A \); it is certainly unimodular. Thus using Lemma 2, the
ordering along the \( a_i \) of those meridional discs making up each \( D_j \) can be
chosen, after inserting any necessary pairs of discs corresponding to \( a_i a_i^{-1} \),
so that, with respect to the new choice, \( \Pi \) becomes a presentation of the
trivial group. Again from Lemma 2, \( \Pi \) is equivalent by Andrews-Curtis
moves to the trivial presentation and so \( M \times I \) is a 5-ball.

References


Received August 15, 2001 and revised October 11, 2001.
University of Cambridge
Department of Pure Mathematics and Mathematical Statistics
Centre for Mathematical Sciences
Wilberforce Road, Cambridge CB3 0WB
UK
E-mail address: wbrl@dpmms.cam.ac.uk
SIMPLIFYING TRIANGULATIONS OF $S^3$

ALEKSANDAR MIJATOVIĆ

In this paper we describe a procedure to simplify any given triangulation of $S^3$ using Pachner moves. We obtain an explicit exponential-type bound on the number of Pachner moves needed for this process. This leads to a new recognition algorithm for the 3-sphere.

1. Introduction.

It has been known for some time that any triangulation of a closed PL $n$-manifold can be transformed into any other triangulation of the same manifold by a finite sequence of moves [5]. We can describe the moves as follows.

**Definition.** Let $T$ be a triangulation of an $n$-manifold $M$. Suppose $D$ is a combinatorial $n$-disc which is a subcomplex both of $T$ and of the boundary of a standard $(n+1)$-simplex $\Delta^{n+1}$. A *Pachner move* consists of changing $T$ by removing the subcomplex $D$ and inserting $\partial\Delta^{n+1} - \text{int}(D)$ (for $n$ equals 3, see Figure 1).

It is an immediate consequence of the definition that there are precisely $(n+1)$ possible Pachner moves in dimension $n$. We can now state Pachner’s result [5] in the following way.

**Theorem 1.1** (Pachner). Closed PL $n$-manifolds $M$ and $N$ with triangulations $T$ and $K$ respectively, are piecewise linearly homeomorphic if and only if there exists a finite sequence of Pachner moves and simplicial isomorphisms taking the triangulation $T$ into the triangulation $K$.

In dimension 3 we have four moves from Figure 1 at our disposal. Using them, we can describe the main theorem of this paper.

**Theorem 1.2.** Let $T$ be a triangulation of a 3-sphere and let $t$ be the number of tetrahedra in it. Then we can simplify the triangulation $T$ to the canonical triangulation of $S^3$, by making less than $a t^2 \frac{2}{3} t$ Pachner moves, where the constant $a$ is bounded above by $6 \cdot 10^6$ and the constant $b$ is smaller than $2 \cdot 10^4$. 

291
Figure 1. Three dimensional Pachner moves.

The triangulation $T$ in this theorem can be non-combinatorial (i.e., simplices are not uniquely determined by their vertices), as is the case with the canonical triangulation of $S^3$, consisting of two standard 3-simplices glued together via an identity on their boundaries. We should mention here that Pachner's original proof of Theorem 1.1 works for combinatorial triangulations only. However, at least in dimension 3, this does not matter because the second derived subdivision of any (possibly non-combinatorial) triangulation is always combinatorial and can be obtained from the original triangulation by a finite sequence of Pachner moves.

A possible effect Pachner's result could have on the theory of 3-manifolds is discussed by the next proposition.

**Proposition 1.3.** Let $T$ and $K$ be two triangulations of the same closed PL 3-manifold $M$. The existence of a computable function, depending only on the number of 3-simplices in $T$ and $K$, bounding the number of Pachner moves required to transform $T$ into $K$, is equivalent to an algorithmic solution of the recognition problem for $M$ among all 3-manifolds.

**Proof.** Assume first that $f(t, k)$ is a computable function as described in the proposition. Suppose that $T$ is a triangulation of $M$ with $t$ 3-simplices. Let $K$ be a triangulation of some closed 3-manifold $N$ containing $k$ 3-simplices. Do all possible sequences of Pachner moves on the triangulation $T$ of length at most $f(t, k)$, and check each time if the result is isomorphic to $K$. This gives an algorithm to determine whether $M$ and $N$ are PL homeomorphic.
Conversely suppose that we have an algorithm to recognize \( M \) among all 3-manifolds. Now we need a complete (finite) list of all triangulations of all 3-manifolds with a fixed number of 3-simplices. In dimension three, such a list can be built algorithmically because there is an easy way of recognizing the 2-sphere (the Euler characteristic suffices) as a link of a vertex.

We can now create all triangulations of \( M \) with the specific number of 3-simplices by running the recognition algorithm for \( M \) (which exists by assumption) on the list of all 3-manifold triangulations with the specified number of 3-simplices.

An algorithm, making all possible Pachner moves on a triangulation of our 3-manifold \( M \) with \( t \) 3-simplices will after a finite number of steps (by Theorem 1.1) necessarily produce a given triangulation of \( M \) containing \( k \) 3-simplices. Since we can list all triangulations of \( M \) with \( t \) (respectively \( k \)) 3-simplices, this gives an algorithm to calculate the value of the function \( f(t, k) \) as required.

At present there is no known algorithm to decide whether a given simplicial complex is an \( n \)-sphere, for \( n \geq 4 \). This means that the proof of one of the implications in Proposition 1.3 breaks down in dimensions five and above since there is no way of building a list of all triangulations of all manifolds with a fixed number of top dimensional simplices, in these dimensions.

The proof of the converse implication in Proposition 1.3, showing that a computable bound implies a recognition algorithm for a given \( n \)-manifold, remains valid in any dimension. Furthermore, if such a computable bound existed for all \( n \)-manifolds, and was independent of the underlying \( n \)-manifold, then it would give an algorithm to determine whether any two \( n \)-manifolds are homeomorphic. But using the fact (proved by A.A. Markov) that there is no such algorithm for \( n \geq 4 \), we can conclude that such a computable function does not exist in dimensions four and above.

It is interesting to note, that for any \( n \)-manifold \( M \) Pachner’s theorem implies the existence of a function, depending only on the number of \( n \)-simplices in \( T \) and \( K \), and bounding the number of Pachner moves necessary for the whole transformation. This is because there are only finitely many triangulations of our \( n \)-manifold \( M \) with fixed numbers of \( n \)-simplices. Then, using Theorem 1.1, a finite sequence of Pachner moves connecting any two of them, can be found. Taking the maximum length over this finite family of sequences gives us the bound. Therefore, computability of the function in Proposition 1.3 is an assumption that can not be omitted.

The upper bound in Theorem 1.2 is computable. It therefore yields a new recognition algorithm for the 3-sphere.

One of the essential ingredients of the proof of Theorem 1.2 is the theory of normal and almost normal surfaces. In this section we shall describe some of its basic features. We will then go on to discuss the Rubinstein-Thompson algorithm \[7\] for recognizing the 3-sphere which provides the setting for the proof of Theorem 1.2. After it, we’ll mention some of the consequences of normal surface theory which will prove to be useful later. At the end of this section we shall prove the isotoping lemma that will later give us a way of simplifying triangulations of the 3-sphere. Let’s start with some definitions.

A normal triangle (respectively quadrilateral) in a 3-simplex $\Delta^3$ is a properly embedded disc $D$, such that its boundary $\partial D$ intersects precisely three (respectively four) edges transversely in a single point and is disjoint from the remaining 1-simplices and vertices of $\Delta^3$. A normal disc is a normal triangle or quadrilateral.

![Figure 2. Three types of normal discs.](image)

There are four possible types of normal triangles, because each triangle is parallel to one of the faces of $\Delta^3$. Normal quadrilaterals will always separate the vertices of the tetrahedron in pairs. It is therefore clear, that we can only have three possible quadrilateral types. Together, there are 7 distinct normal disc types in a tetrahedron.

Let $M$ be a 3-manifold with a triangulation $T$. A properly embedded surface $F$ in $M$ is in normal form with respect to the triangulation $T$, if it intersects each tetrahedron of $T$ in a finite (possibly empty) collection of disjoint normal discs.

Since normal surfaces are always embedded, at most one of the quadrilateral types can occur in each 3-simplex.

Suppose $F$ is a normal surface in $M$ with respect to $T$. Then $F$ corresponds to a vector $\mathbf{x} = (x_1, \ldots, x_{7t})$ with $7t$ coordinates, where $t$ denotes the number of 3-simplices in the triangulation $T$. The index set $\{1, \ldots, 7t\}$ corresponds to all possible disc types in $T$ (there is 7 of them for each tetrahedron). The coordinate $x_i$ is simply the number of copies of $i$-th disc type in our surface $F$. 
Each 2-simplex in $T$ contains three types of normal arcs (coming from normal discs), one cutting off each vertex of the triangle. If it is a face of two 3-simplices in $T$, then it gives rise to three matching (linear) equations, one corresponding to each normal arc type. Doing this for every triangle, not in the boundary of $M$, we’ve constructed a linear system in $7t$ variables, consisting of at most $6t$ equations.

It follows immediately from the construction, that the vector $x$, coming from the normal surface $F$, gives a solution to the linear system. By imposing extra conditions to ensure that all quadrilaterals in a given tetrahedron are of the same type, we obtain a restricted linear system. The conditions we’ve just added are sometimes referred to as quadrilateral constraints. Now there is a one to one correspondence between embedded normal surfaces in $M$ and nonnegative integral solutions to the restricted linear system.

Haken proved that all nonnegative integral solutions to such a system are integer linear combinations of a finite set of nonnegative integral solutions $x_1, \ldots, x_n$, called fundamental solutions, which can be found in an algorithmic way. As it turns out, these fundamental solutions are characterized by the property of not having a decomposition as a sum of two (nontrivial) nonnegative integral solutions to the restricted linear system.

Since each fundamental solution corresponds to an embedded normal surface, we obtain a finite set $F_1, \ldots, F_n$ of embedded normal surfaces, called fundamental surfaces. Any embedded normal surface in $M$ can thus be written algebraically as a nonnegative integer linear combination of fundamental surfaces. Miraculously, this algebraic fact carries over to geometry. In other words, we can define a geometric addition for any two normal surfaces $F$ and $G$ with the property that the sum of the corresponding solutions to the restricted linear system, is again a solution of the same system. This condition boils down to the fact that the union of all normal discs in both $F$ and $G$ satisfies the quadrilateral constraints.

Assuming that and putting both surfaces in general position with respect to one another, cutting along the arcs of intersection in each tetrahedron, and pasting the pieces back together in the unique way, so that we end up with normal discs only, yields a well-defined embedded normal surface $F + G$. Its corresponding vector is a sum of the vectors coming from $F$ and $G$. The cut and paste process described above is sometimes called regular alteration.

An isotopy of the ambient manifold, preserving the normal structure of a given normal surface is called a normal isotopy. We should also note that the geometric addition described above is well-defined up to a normal isotopy of the summands.

Before we describe the Rubinstein-Thompson algorithm, we need to introduce a concept, originally due to Rubinstein.
Definition. A properly embedded surface in a 3-manifold $M$ with a triangulation $T$ is almost normal with respect to $T$, if it intersects each tetrahedron of $T$ in a finite (possibly empty) collection of disjoint normal discs except in precisely one tetrahedron there is precisely one exceptional piece from Figure 3 and possibly some normal triangles.

![Almost normal pieces](image)

Figure 3. Almost normal pieces.

This exceptional piece is either a disc (the first possibility in Figure 3) whose boundary is a normal curve of length eight (i.e., an octagon), or it is an annulus consisting of two normal disc types with a tube between them that is parallel to an edge of the 1-skeleton.

Now we can describe the Rubinstein-Thompson algorithm which is designed to determine whether or not a 3-manifold $M$ with a triangulation $T$ is a 3-sphere. We can assume that $M$ is closed, orientable and that $H_1(M;\mathbb{Z}_2)$ is trivial. All these properties can be checked algorithmically. The last assumption guarantees that $M$ contains no closed non-separating surfaces. The algorithm now is in three steps. We proceed as follows.

Step 1. Find a maximal collection $\Sigma$ of disjoint non-parallel normal 2-spheres in $M$.

Step 2. Cut $M$ open along $\Sigma$. This splits $M$ into three different types of pieces:

Type A: A 3-ball neighborhood of a vertex of $T$ (every vertex is enclosed in such a piece).

Type B: A piece with more than one boundary component.

Type C: A piece with exactly one boundary component which is not of Type A.
Step 3. Search each Type C piece for an almost normal 2-sphere with an octagonal component.

Conclusion: $M$ is a 3-sphere if and only if every Type C piece contains an almost normal 2-sphere with an octagonal component.

The bulk of the proof that this indeed is a recognition algorithm for the 3-sphere relies on the following two lemmas from [7].

**Lemma 2.1.** A Type B piece is a punctured 3-ball.

**Lemma 2.2.** A Type C piece is a 3-ball if and only if it contains an "octagonal" almost normal 2-sphere.

By Lemma 2.2, if some Type C piece fails to contain an “octagonal” almost normal 2-sphere, then it is not a 3-ball and $M$ is not a 3-sphere. Otherwise, $M$ is just a collection of 3-balls and punctured 3-balls glued together. Since every 2-sphere is separating, $M$ has to be a 3-sphere.

The difficult part of the argument is in the proof of Lemma 2.2. It is here that Thompson simplified Rubinstein’s original methods to prove the existence of an “octagonal” almost normal 2-sphere in a 3-ball of Type C, by using Gabai’s powerful notion of thin position. We should also note that the easier converse implication in Lemma 2.2 follows from Lemma 2.7.

In order to be able find a maximal collection of disjoint non-parallel normal 2-spheres in $M$ in an algorithmic way, we need the following lemma.

**Lemma 2.3.** A maximal collection $\Sigma$ of disjoint non-parallel normal 2-spheres in $M$, as in the Rubinstein-Thompson algorithm, can always be constructed algorithmically.

A proof of this lemma was given by Casson [1]. It was also described in [3] (see Lemma 3). It will be important for us to be able to bound the complexity of all of the 2-spheres in the maximal collection $\Sigma$. We shall therefore give a brief description of this algorithm.

Additivity of Euler characteristic implies at once that if there exists a non-trivial normal 2-sphere in our triangulation, we can also find one (which is also nontrivial) among fundamental surfaces. Since the family of fundamental surfaces is accessible in an algorithmic way, we can take this fundamental 2-sphere to be the first element in $\Sigma$.

Assume now that we have already constructed a subcollection $\Sigma'$ of $\Sigma$. We shall look for normal surfaces with respect to the triangulation $T$, lying in the complement of the normal 2-spheres constructed so far. In any tetrahedron from $T$ we can have complementary regions of $\Sigma'$ that are not of the form triangle $\times I$ or square $\times I$ (see Figure 4), as well as the ones that are. Note also that the unions of the product regions support a natural $I$-bundle structure. These $I$-bundles are usually referred to as parallelity regions.

We can now describe normal surfaces in the complement of $\Sigma'$ by assigning a variable to each triangle or square type that does not lie in any of the
parallelity regions. The equations are again the matching equations along
the faces together with the equations that ensure that the surface intersects
each parallelity region in a well-defined number of components. In other
words each parallelity \( I \)-bundle contributes new linear equations (that are
of the same form as the matching equations along faces) and no new vari-
ables. This is because the normal 2-spheres we are looking for, can only run
parallel to the horizontal boundary in these \( I \)-bundles. Adding the usual
quadrilateral constraints gives a restricted linear system.

Like in standard normal surface theory, we have a one to one correspon-
dence between closed normal surfaces in the complement of \( \Sigma' \) and solutions
of the above restricted linear system. Since addition of two solutions is
again realized topologically by regular alteration, the same argument as be-
fore tells us that the next normal 2-sphere, that is not parallel to any of the
elements in \( \Sigma' \), can be chosen from the family of fundamental surfaces.

So in order to find a maximal family \( \Sigma \) of disjoint non-parallel normal 2-
spheres, we just have to keep repeating this procedure. We stop when each
normal 2-sphere in the complement of \( \Sigma' \) is normally parallel to some normal
2-spheres in the collection \( \Sigma' \). Lemma 2.4 guarantees that this process has
to reach such a stage.

As far as the complexity, i.e., the number of normal pieces, of elements
in \( \Sigma \) goes, at each stage it is going to be bounded by Proposition 2.5. Since
the linear algebra in the proof of Proposition 2.5 (which can be found in [2])
depends only on the number of normal variables and is independent of how
many equations we have, it is the number of different normal disc types
outside the parallelity regions, that needs to be controlled. The proof of
Lemma 2.4 shows that this number is bounded linearly by the number of
tetrahedra in \( T \). In fact there can be at most 11 different normal disc types
in a single tetrahedron in \( T \) at any stage of the process. We shall calculate
explicit upper bounds later on in this section.

We still need to answer the question of how to search for “octagonal”
almost normal 2-spheres that are contained in Type C pieces. Modified
versions of standard normal surface theory algorithms suffice for the search.
So our goal is to construct an algorithmic procedure which will find an
“octagonal” almost normal 2-sphere in each Type C piece. These 2-spheres
will exists by Lemma 2.2 if the 3-manifold \( M \) we are looking at is a 3-sphere.
We proceed as follows.

First we fix a tetrahedron \( H \) in the triangulation \( T \) of a 3-manifold \( M \)
and then we fix a normal curve \( c \) of length eight on its boundary (there
are three choices for \( c \)). Now an analogue to the normal surface theory,
used to construct the collection \( \Sigma \), can be set up. The matching conditions
will look just like before. Quadrilateral constraints have to be modified
however, because we want our solutions to consist of normal triangles and
quadrilaterals everywhere except in \( H \), where we want them to be composed
of normal triangles and octagonal components with boundaries parallel to $c$. The notion of regular alteration can be defined in this generalized setting and again it gives rise to the correspondence between integer linear combinations of the fundamental solutions to the (generalized) restricted linear system and the set of all surfaces described above. Fundamental surfaces are again the ones corresponding to fundamental solutions. We should also note that their complexity is bounded by Proposition 2.5 since the linear system they are the solutions of, has less than $11t$ variables.

What we really want is to find algorithmically “octagonal” almost normal 2-spheres that are contained in Type C pieces. We know that one such 2-sphere exists in each Type C piece by Lemma 2.2. This 2-sphere can be expressed as a sum of the fundamental surfaces. Precisely one of the summands has to contain a single octagonal piece and, since the Euler characteristic is additive, at least one of the fundamental surfaces in the sum has to be a 2-sphere (since the Type C piece we are looking at contains an “octagonal” almost normal 2-sphere, it has to be a 3-ball and can therefore not contain embedded projective planes). If the fundamental 2-sphere in the sum does not contain an octagon, then it has to be normal and thus parallel to the unique normal 2-sphere from $\Sigma$ that is bounding the Type C piece we are looking at. This is a contradiction because we could then isotope it away from all the other summands by a normal isotopy. Since regular alteration is defined up to normal isotopy, this would then make the sum (i.e., a 2-sphere) disconnected. So we’ve found an “octagonal” almost normal 2-sphere in a Type C piece that is fundamental.

The complexity of the fundamental “octagonal” almost normal 2-sphere we’ve just constructed is bounded in the same way as all the other complexities of the normal 2-spheres in $\Sigma$. This follows directly from the construction, since all we are doing when searching for an almost normal 2-sphere, is just making another step of the recursion that gave us $\Sigma$, without increasing the number of normal variables. We will give an explicit estimate for the complexity later on in this section.

Let’s first bound the number of disjoint non-parallel normal 2-spheres in $\Sigma$. This is made possible by an old idea due to Kneser.

**Lemma 2.4.** Let $T$ be any triangulation of $S^3$ and let $t$ be the number of tetrahedra in $T$. Then any family of disjoint non-parallel normal 2-spheres contains at most $6t$ of them.

**Proof.** Normal triangles and squares chop up any tetrahedron in $T$ into several pieces. But at most six of these regions are not of the form triangle $\times I$ or square $\times I$ (see Figure 4).

Let $n$ be the maximal number of disjoint non-parallel normal 2-spheres in $T$. Then the complement of this family has precisely $(n + 1)$ components. Each of those components must contain at least one of the non-product
regions. This is because any component, consisting only of product pieces, is bounded by two parallel normal 2-spheres. Since the total number of non-product regions is bounded by $6t$, our lemma is proved.

We are interested in bounding the number of normal pieces of elements in $\Sigma$. We also want to bound the number of normal pieces of the “octagonal” almost normal 2-spheres that arise in Type C pieces. Both of these things can be accomplished at one go, because we know that the procedure giving $\Sigma$ can be extended (by making a single additional step) to an algorithm producing “octagonal” almost normal 2-spheres in Type C pieces.

The proposition we are about to state is proved in \[2\]. It originally deals with the linear system in $7t$ variables coming from the matching equations for normal surfaces. Its proof uses some basic linear algebra on the linear system which consists of matching equations. We should note at this point that the number of equations in this linear system does not influence the bound that the proposition gives.

**Proposition 2.5.** Let $M$ be a triangulated 3-manifold containing $t$ tetrahedra. Let $x$ be a fundamental solution of a system of linear equations coming from matching conditions. Then each coordinate of the vector $x$ is bounded above by $7t^2t^{-1}$.

Using Proposition 2.5, we can bound the size of each component of all the vectors corresponding to the normal 2-spheres in $\Sigma$. It follows from Figure 4 that the number of normal discs that are not contained in the parallelity regions (at any stage of the algorithm producing the family $\Sigma$) is always bounded above by $11t$. The system of equations we are solving at each stage consists of the matching equations along the faces together with the equations that ensure that the surface intersects each parallelity region in a well-defined number of components. We should note here that the latter equations are of the same form as the matching equations.
Proposition 2.5 then implies that there can be at most $11t2^{11t-1}$ parallel copies of a given normal disc type in a complement of any subcollection $\Sigma'$ from any stage of the algorithm. This can be deduced because the proof of Proposition 2.5 depends only on the number of variables and the shape of the equations, i.e., the number and size of nonzero coefficients in each equation. Since the number of variables has increased and the shape of the equations hasn’t changed, we get the above bound by substituting $7t$ with $11t$ into Proposition 2.5.

Lemma 2.4 tells us that we’ll never have to make more than $6t$ steps when constructing $\Sigma$. This means that each of the normal disc types in the complement of any subcollection $\Sigma'$ can only give rise to less than $(2 \cdot 11t \cdot 2^{11t-1})^6t$ normal discs of the same type in the initial triangulation $T$. This is because at each stage of the algorithm the number of parallel copies of a fixed normal disc type is given by $11t2^{11t-1}$. We have to include the factor of 2 because in each parallelity region there are two normal variables contributing to the number of parallel copies of a given normal disc type in the initial triangulation.

We can obtain a similar kind of bound for “octagonal” almost normal 2-spheres. We only have to change the exponent from $6t$ to $(6t + 1)$. This is because all these “octagonal” almost normal 2-spheres are just one step away (in our algorithm) from the normal ones (bounding Type C pieces) and at each stage they are described by fewer variables. For example, in our original triangulation they require $7(t - 1) + 4$ variables. So the bound in Proposition 2.5 applies.

Putting everything together and using the fact that $5t(11t2^{11t})^{6t+1} < 2^{110t^2}$, we get the following lemma (the factor $5t$ comes in because there are at most 5 different normal and almost normal pieces in each tetrahedron of $T$).

**Lemma 2.6.** Let $T$ be a triangulation of the 3-sphere which contains $t$ tetrahedra. Then the number of all normal pieces contained both in all elements of $\Sigma$ and in all “octagonal” almost normal 2-spheres from all Type C pieces is bounded above by $2^{110t^2}$.

We should note at this point that this is the only part of the bound in Theorem 1.2 which contains a quadratic expression in its exponent. If one could find both the “octagonal” almost normal 2-spheres in Type C pieces and the maximal family $\Sigma$ among the fundamental solutions of linear systems that are based on the triangulation $T$, the bound in Lemma 2.6 would have a linear function (similar to the one in Proposition 2.5) in its exponent.

The essential process we are just about to describe, is the one of isotoping almost normal surfaces. It is going to provide a foundation for the simplifying procedure needed for the proof of Theorem 1.2.
Let $F$ be a separating almost normal surface in a 3-manifold $M$ with a triangulation $T$. Its weight, $w(F)$, is defined to be the number of points in the intersection of $F$ and the 1-skeleton $T^1$. If $F$ contains an octagon, a natural isotopy is the one pushing the surface over an edge which meets the length eight normal curve bounding the octagon in two points. There are two possible natural isotopies, depending on the component of $M - F$ we are pushing into. In case of other non-normal pieces (see Figure 3), a natural isotopy pushes the tube part of the annulus so that it encompasses one of the edges it is parallel to. As a result in both cases, we get a surface with its weight equal to $w(F) - 2$.

Notice that if we look at our almost normal surface $F$ in the complement of the 1-skeleton of $T$, there is an obvious compression disc $D$ for it, enveloping the edge we are isotoping over. The natural isotopy can then be realized by isotoping over the 3-ball bounded by $D$ and the disc in $F$, bounded by $\partial D$.

The natural isotopy is only the first step in the process of isotoping almost normal surfaces. Everything else will be accomplished by a sequence of elementary isotopies. We can define them as follows.

Let $A$ be a 2-simplex in $T$, containing a non-normal arc (Figure 5) which comes from intersecting $A$ with an isotope of $F$.

![Figure 5](image)

**Figure 5.** An isotope of $F$ intersects $A$ in a non-normal arc.

A disc $B$ in the triangle $A$ (see Figure 5) is bounded by the non-normal arc and a subarc of the edge $e$. An elementary disc can be constructed by banding together two parallel copies of $B$ in the complement of 1-skeleton, where the band runs around the edge $e$. Its boundary is a simple closed curve in the surface, bounding a disc on one side. An elementary isotopy is an isotopy over the 3-ball bounded by the disc in the isotope of $F$ and the elementary disc we’ve just defined.

Since $F$ is a separating surface, we can fix a complementary component $I$ of $M - F$. All the elementary isotopies that we are going to do from now on, are going to have the same direction. We will always be isotoping towards the interior of the component $I$.

The following isotoping lemma will play a crucial role in the simplifying process. A similar result is proved in [6] by a careful inspection of all the possible cases. The proof we are giving here is based on elementary isotopies.
and is better suited from our perspective because it sheds more light on the side of things we'll be interested in later.

**Lemma 2.7.** Let $F$ be a separating almost normal surface in a 3-manifold $M$ with a triangulation $T$. Let $I$ be a component of the complement $M - F$, if $F$ contains an octagon. Otherwise let $I$ denote the component containing a solid torus region in the interior of the 3-simplex where $F$ is not normal. A natural isotopy followed by a sequence of all possible elementary isotopies, all going in the direction of $I$, will result with a surface intersecting each tetrahedron of $T$ in pieces as in Figure 6 and in normal pieces. Moreover, in each tetrahedron there can only be at most one piece of the first type from Figure 6. A single 3-simplex can contain several pieces of all the other types in Figure 6 as well as several normal pieces. The pieces in Figure 6 can not be parallel.

![Figure 6. Non-normal pieces in the tetrahedra of $T$.](image)

**Proof.** First note that after the natural isotopy, all the non-normal arcs we get will give rise to elementary isotopies in the direction of $I$. After each isotopy both $F$ and $I$ will change, but we'll still denote both resulting spaces by $F$ and $I$ respectively.

After the natural isotopy, $F$ and $I$ satisfy the following conditions:

1) In each tetrahedron of $T$ the component $I$ consists of a family of 3-balls, each one bounded by pieces of $F$ and a (possibly disconnected) planar surface, contained in the boundary of the 3-simplex.

2) Each 3-ball from 1 intersects any face of the tetrahedron it lies in, in at most one disc.
An elementary isotopy moves a disc in $F$ over a 3-ball in $I$, which intersects a single edge $e$ in $T^1$. So the new $I$ is just the old $I$ without the 3-ball we isotoped over. This 3-ball is a union of a family of 3-balls, one in each tetrahedron of the star of $e$. In fact there can be more than one 3-ball from the same family in a single 3-simplex if this 3-simplex occurs more than once in the star of the edge $e$. This is perfectly feasible in a non-combinatorial triangulation, but it does not have any effect on the process we are studying.

The elements of the above family are the ones that are going to determine the topology of the pieces of $I$ in tetrahedra of $T$. In fact, each 3-ball from condition 1 will after an elementary isotopy still satisfy both conditions if we substitute the old $F$ with the new one. So after performing all possible elementary isotopies towards the interior of $I$, the surface $F$ we end up with will intersect each triangle of $T$ in normal arcs and simple closed curves which miss the boundary of the triangle.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure7.png}
\caption{An intermediate state of the isotopy on a triangle in $T$.}
\end{figure}

The region $I$ will, after the isotopy, consist of 3-balls in each 3-simplex. There is going to be a bijective correspondence between the 3-balls in the end, and the ones we started with. By condition 2, every 3-ball will still intersect any face of the 3-simplex it lies in, in at most one disc. It is also true that the number of these discs will not increase when we pass from the 3-ball pieces of $I$ at the beginning to the 3-balls at the end.

Let's look at the pieces of $F$ in each tetrahedron. It is obvious that all the possibilities of the lemma can actually arise. We have to see that they are the only ones.

**Claim.** A single piece of $F$ can intersect a triangle of $T$ in either a unique normal arc or in a single simple closed curve.

Every piece of $F$ is contained in the boundary of a 3-ball piece of $I$. This 3-ball intersects each triangle of the 2-skeleton $T^2$ in at most one disc. So no triangle can contain two simple closed curves or a simple closed curve and a normal arc, both belonging to the same piece of $F$.

The same argument tells us that a triangle in $T^2$ can either contain two normal arcs of intersection with a single piece of $F$ or at most three of them, each one cutting off a vertex of the triangle in the 2-skeleton.
Now we need to prove that our piece of $F$ can have at most one normal simple closed curve boundary component. So assume the opposite. Since the piece is a subset of the boundary of a 3-ball, no arc contained in it, running between two distinct boundary components of our piece, can be extended to a simple closed curve in the 2-sphere bounding that 3-ball, without increasing the number of intersection points with the boundary of our piece. On the other hand, assuming we have at least two normal simple closed curves in the boundary, there surely exist two normal arcs, belonging to the distinct boundary components of our piece, that are contained in a single 2-simplex. Connecting them by an arc in the piece of $F$ contradicts what was said before (because these two normal arcs are both contained in the boundary of a disc in the 2-simplex they lie in).

So now it follows that the piece of $F$ we are looking at, can contain at most one normal boundary component which is of length at most eight. This is because the only normal curve of length 12, intersecting each 2-simplex (in the boundary of a tetrahedron) in 3 normal arcs as above, consists of 4 simple closed curves (one for each vertex of the tetrahedron). It is also well-known that normal simple closed curves of lengths 9, 10 or 11 do not exist.

There are precisely three normal simple closed curves of length eight in the boundary of a tetrahedron. So if our piece of $F$ is bounded by one such curve, then at least one of the faces of the tetrahedron intersects the 3-ball piece of $I$ (containing in its boundary the piece of $F$ we are considering) in two discs. This is a contradiction that proves the claim.

The claim implies the following seven possible boundaries for any piece of $F$: Normal simple closed curve of length three, normal simple closed curve of length four, single simple closed curve, normal simple closed curve of length three and a simple closed curve, two simple closed curves, three simple closed curves, four simple closed curves.

Since every piece of $F$ in any tetrahedron is planar, it is up to homeomorphism determined by its boundary. This implies that all possible pieces of $F$ are the ones listed in the lemma.

The fact that all these planar surfaces are embedded as in Figure 6 (up to an isomorphism of the tetrahedron) follows from the observation that all the elementary discs are parallel to edges of the 1-skeleton. □

3. Outline of the proof.

Given a triangulation $T$ of the 3-sphere, how do we simplify it? The process is divided into two stages. First, we create a subdivision $S$ of $T$ by defining it in each complementary piece of the manifold $S^3 - \Sigma$ in such a way that the triangulations match along all normal 2-spheres in $\Sigma$. The second step consists of simplifying $S$ down to the canonical triangulation of $S^3$. 
An explicit construction of $S$, using Pachner moves, will be given in Section 5. The simplifying procedure of step two is based on the relationship between Pachner moves and shellable triangulations. This relationship will be established in Section 4.

Now, we are going to describe the additional structure on the complementary pieces of $S^3 - \Sigma$, needed for the definition of the subdivision $S$. We already know (Lemma 2.2) that every Type C piece contains an “octagonal” almost normal 2-sphere. To see that each Type B piece also contains an almost normal 2-sphere, it is useful to introduce an ordering on the normal family $\Sigma$. It comes naturally by picking a vertex of $T$ and looking at the complementary region (which is not a Type A piece) of the trivial normal 2-sphere around it. Topologically we get a 3-ball containing our normal family $\Sigma$. Now, the ordering on $\Sigma$ is induced by inclusion. For example, the trivial normal 2-sphere around the vertex we removed is the unique largest element. The smallest elements in this ordering are either trivial normal 2-spheres consisting of normal triangles only or the ones bounding Type C 3-balls.

Our task is to find an almost normal 2-sphere in a piece with more than one boundary component. Pick the largest 2-sphere in its boundary. A very nice argument in [7] (subclaim 2.0.1.) implies that there must be an edge in $T$ with a subarc which runs from the largest component of the boundary to some other component and whose interior is disjoint from $\Sigma$. By taking parallel copies of the two 2-spheres connected by this arc in the piece we are looking at, and tubing them together in one of the tetrahedra in the star of the edge, we obtain our almost normal 2-sphere.

All the almost normal 2-spheres we’ve created are separating, because we are in $S^3$. By picking the right complementary component in $S^3$ and applying Lemma 2.7, we can simplify each almost normal 2-sphere by a sequence of elementary isotopies. Since we are only using elementary isotopies going in the same direction (towards the interior of a fixed complementary component in $S^3$), the whole process can be realized by an embedding of $S^2 \times I$, where the top 2-sphere is the almost normal 2-sphere we started with and the bottom one is the 2-sphere coming from Lemma 2.7 (see Figures 8 and 9).

Another important point here is that the whole isotopy never leaves the Type B (or C) piece it started in. This is true simply because an analogous statement holds for each elementary isotopy. This implies that the isotoped surface coming from Lemma 2.7 will be contained in the interior of the piece containing the almost normal 2-sphere we started with.

In a Type C piece, the isotopy can go in two directions because the almost normal 2-sphere in this case contains an octagon. The surface we get, when isotoping towards the interior of the piece, will have 0 weight. This follows from the observation that we can forget about all pieces in Figure 6 if we
compress each annulus with a length three normal curve in its boundary. This would then give a family of normal 2-spheres contained in the Type C piece which is a contradiction. Therefore, the 2-sphere we end up with has to miss the 1-skeleton.

Similar reasoning tells us that an isotopy in the other direction in the Type C piece has to end with a 2-sphere, intersecting the 2-skeleton $T^2$ in normal curves parallel to the ones coming from the boundary of the piece we are looking at and possibly in some simple closed curves which miss the 1-skeleton.

The almost normal 2-spheres in Type B pieces that we are going to consider, will never contain an octagon. We will therefore be isotoping in one direction only. Using the same kind of arguments as before, we can conclude that the 2-sphere we end up with consists of boundary components of the Type B piece (all except the two we started with, see Figure 9), tubed together by pieces depicted in Figure 6. It should be noted that if a Type B piece has only two boundary components, then the isotoped 2-sphere does not intersect the 1-skeleton.

Let $\Lambda$ be the following collection of 2-spheres: In every Type C piece just take an “octagonal” almost normal 2-sphere which exists by Lemma 2.2. In each Type B piece take a copy of the almost normal 2-sphere described above with the annulus connecting two normal pieces moved by a natural isotopy, so that it envelops the edge it is parallel to. The 2-spheres from $\Lambda$ in Type B pieces are therefore normal in all the tetrahedra of $T$, except in the ones contained in the star of the edge we isotoped over.

The sequences of elementary 3-balls, corresponding to the supports of elementary isotopies, yields the additional structure (on Type B and C pieces) that is required to define the subdivision $S$. Elementary discs and an element in $\Lambda$ chop up each Type C piece of $S^3 - \Sigma$ (see Figure 8).

In the case of a Type B piece, the element of $\Lambda$ will, after the isotopy, consist of all but two boundary normal 2-spheres tubed together by pieces described in Figure 6. Again, the Type B piece in question can be decomposed into (many) 3-balls and two punctured 3-balls. The two punctured 3-balls come from an element in $\Lambda$ we started our isotopy on, and from its isotope after we’ve performed all elementary isotopies on the “tubed” almost normal 2-sphere (see Figure 9). The rest of the Type B piece is decomposed into 3-balls by all the elementary discs required for this isotopy.

After the isotopy from Lemma 2.7, what’s left in each tetrahedron of the complementary component we isotoped into, are just 3-balls bounded by the pieces from Lemma 2.7 on one side and possibly some normal pieces of the elements in $\Sigma$ on the other. Schematically, the situation after the isotopy is depicted by Figure 9.

Now we want to triangulate all of these 3-balls (the elementary ones as well as the ones that are left over in the component we were isotoping into),
in all the pieces of the complement of $\Sigma$, by simple shellable triangulations. Since all the processes described above induce polyhedral structures in the boundaries of all the 3-balls (this will be described in detail in Section 5) in question, subdividing the boundary 2-spheres in order to obtain genuine triangulations and then coning them, does the job. Doing so in every piece of the space $S^3 - \Sigma$ exhausts the whole 3-sphere and therefore completely determines the subdivision $S$.

The fact that all these cones are indeed shellable, is proved in [4] (Lemma 5.4). Here we are relying on the property that all the bounding 2-spheres we’ll need to cone in the process, are triangulated by combinatorial triangulations. The reason why we want these 3-balls to be shellable is simply

**Figure 8.** Two sequences of elementary isotopies in a Type C piece.

**Figure 9.** The “tubed” almost normal 2-sphere and the isotoped 2-sphere in a Type B piece.
because each elementary isotopy can then be realized by a shelling of the corresponding 3-ball.

So what we really want from the cones on the 2-spheres above, is to be shsemblable without ever having to shell from the faces contained in a fixed disc, which lies in the bounding 2-sphere. This disc is just a 2-manifold along which the 3-ball, we are trying to triangulate, is glued onto the rest of the (Type B or C) piece. This can always be achieved since a cone on a disc with a combinatorial triangulation can be shelled “from the side” just by coning the shelling procedure of the disc itself.

The simplifying process works its way up the ordering of the normal 2-spheres in $\Sigma$. First we change the subdivision $S$ in each Type C piece (which are smallest elements in our ordering), making it a cone on the unique boundary component. In Section 4 we will discuss how to implement elementary shellings from a 2-sphere boundary component by Pachner moves, if on the other side of that 2-sphere we have a cone on it. Using that construction we can pick a 3-ball piece (in some 3-simplex), contained in the 3-ball $X$ from Figure 8, and turn the whole 3-ball $X$, bounded by the 2-sphere coming out of Lemma 2.7, into a cone on its boundary. This is simply because the complement (in $X$) of the coned 3-ball piece we picked, is shsemblable. That follows from the observation that all the (coned) 3-ball regions from Figure 6, the 3-ball $X$ is made of (we already know that the first possibility in Figure 6 can not occur) can be viewed as vertices of a graph whose edges correspond to the discs in the interiors of the 2-simplices of $T$. Since this graph is a tree (this follows from the fact that the isotoped 2-sphere bounds a 3-ball), there is a “global” shelling strategy for the complement of the piece we picked in the 3-ball $X$. This can be made simplicial by shelling one cone at a time.

So now we can assume that the 3-ball $X$ is coned. We can carry on by shelling (in the reversed order) all elementary 3-balls (and the 3-ball corresponding to the natural isotopy) involved in the isotopy taking the almost normal 2-sphere $Y$ to the boundary of $X$. By this stage, we’ve changed the subdivision $S$ so that it looks like a cone on $Y$ in the Type C piece we are looking at. Above it, $S$ is still unchanged. We can now do the same thing towards the boundary of the piece we are considering, again using the shsemblable nature of the subdivision $S$ in all appropriate 3-balls.

What we have now is a cone on the isotope of $Y$. Now we concentrate on the $S^2 \times I$ region between the isotope of the “octagonal” almost normal 2-sphere $Y$ and the single boundary component of the Type C piece. We first shell all the 3-balls from that region that are bounded by pieces in Figure 6. We thus obtain a cone on a normal 2-sphere which is parallel to the bounding 2-sphere of our piece. Since all the regions between any two parallel normal pieces are cones as well, we can shell them one by one and therefore get a cone on the boundary of our piece. Here we are relying on the fact that the normal structure on the bounding 2-sphere is shsemblable. In
general this needn’t hold, but the technical assumption that we are going to make on our triangulation $T$ at the beginning of Section 5 will guarantee this property. This completes the simplification of the triangulation $S$ in all Type C pieces.

Take a Type B piece and assume that all normal 2-spheres, strictly smaller than the largest normal 2-sphere in its boundary, already bound coned 3-balls. The strategy now is similar to the one we used in Type C pieces. Since the tube of the 2-sphere element of $\Lambda$ in our piece runs from the largest boundary component to some other boundary component, we can deduce that all other normal 2-spheres in the boundary that are going to be tubed together by the isotoped 2-sphere (see Figure 9), are going to bound cones on one side.

We first shell all the regions which are bounded by two parallel normal pieces and lie between a normal boundary component and the isotoped 2-sphere. We can do that by expanding the cone structures on the other side of the boundary components of our Type B piece, that exist by assumption. Now the 3-ball bounded by the isotoped 2-sphere from $\Lambda$ is again chopped up into 3-ball regions that are glued together along discs contained in the interiors of 2-simplices of $T$. Like before, there is a sequence of elementary shellings which gives a way of changing the triangulation of the 3-ball bounded by the isotoped 2-sphere from $\Lambda$ to the cone on its boundary.

We will now mimic what we did in Type C pieces. Let’s take the sequence (in the reversed order) of all elementary 3-balls coming from the elementary isotopies needed to push the 2-sphere element of $\Lambda$ in our Type B piece, down to the 2-sphere which now already bounds a coned 3-ball. Using this sequence in the same way as above, we can change the triangulation $S$ in our piece to the cone on the 2-sphere element of $\Lambda$. It is now obvious how to simplify the remains of the subdivision $S$ in the Type B piece we are looking at. So we’ve managed to transform the subdivision in our piece into a cone on the largest boundary component.

Now we want to make sure that the techniques described above suffice for the total simplification of the subdivision $S$. Let’s assume that $K$ is a triangulation of $S^3$ and that $S$ is its subdivision containing $\Sigma$ as a subcomplex. Let $x$ be the vertex of $K$ inducing an ordering on $\Sigma$. Let’s also assume the following property: If all elements of $\Sigma$, smaller than a given normal 2-sphere $A$ in $\Sigma$, bound coned 3-balls, then using Pachner moves we can change the triangulation of the 3-ball bounded by $A$ into a cone on $A$, without altering the simplicial structure of $A$. Then we claim that we can transform $S$, using Pachner moves only, into a cone on $x$ glued to another copy of itself via an identity on the boundary.

To see this, we’ll use a simple induction on the depth of elements in $\Sigma$. A normal 2-sphere in $\Sigma$ is of depth $k$ if it is greater than precisely $k$ elements of $\Sigma$. 
We can use our assumption for the 2-spheres of depth 0. Let $A$ in $\Sigma$ be of depth $(k + 1)$ and assume we’ve coned all the 2-spheres of depth smaller or equal to $k$. Any 2-sphere smaller than $A$ is of depth at most $k$. So we can use the assumption again. This proves our claim and therefore completes the simplification process.

4. Pachner moves and shellable triangulations.

In this section we are going to establish a relationship between elementary shellings and Pachner moves. We will do this in dimensions two and three. Both cases will play a crucial role in building and simplifying the subdivided triangulation $S$. Let’s start by stating precisely what we mean by shelling.

**Definition.** Suppose that $M'$ is a submanifold of a triangulated $n$-manifold $M$ with boundary. If there exists an $n$-simplex $\Delta$ in the triangulation of $M$ with the property that $\Delta \cap \partial M$ is a combinatorial $(n - 1)$-disc, such that $M'$ equals the closure (in $M$) of the complement $M - \Delta$, then we say that $M'$ is obtained from $M$ by an *elementary shelling*.

An elementary shelling is quite similar to an elementary collapse of the top dimensional simplex. The crucial difference lies in the fact that here we stipulate explicitly that the resulting space has to be a manifold.

Another thing which is worth mentioning is that the boundaries $\partial M$ and $\partial M'$ differ by a single $n - 1$ dimensional Pachner move.

A sequence of such elementary shellings is called a *shelling*. Saying that a triangulation of an $n$-manifold is *shellable* simply means that there exists a sequence of elementary shellings which will reduce the triangulation down to a single $n$-simplex. Since the homeomorphism type of the manifold in question does not change under an elementary shelling, it is clear that $n$-balls are the only candidates to have shellable triangulations. It is for example very well-known that any combinatorial triangulation of the two dimensional disc is always shellable. As it was mentioned before, Lemma 5.4 in [4] and the above observation about discs together imply that a cone on any combinatorial triangulation of the 2-sphere constitutes a shellable triangulation of the 3-ball.

Now we are going to express all possible elementary shellings by Pachner moves in the following three dimensional situation. Suppose we had a triangulated 3-manifold and we wanted to make an elementary shelling from a 2-sphere boundary component. Suppose further that on the other side of this 2-sphere, we had a cone on it. We have to consider three different cases according to the number of faces of the 3-simplex we are shelling, which are contained in the boundary 2-sphere.

The first case, where we have a single triangle in the boundary, is dealt with by Figure 10.
Figure 10. A single free face requires one (2-3) Pachner move.

We should note that before making the (2-3) move in Figure 10, the top 3-simplex is contained in the manifold, while the bottom one belongs to the cone. After the move, all three 3-simplices are contained in the altered cone.

The second case is the one where we have two faces in the boundary. It is clear from Figure 11, that a single (3-2) Pachner move suffices.

Figure 11. A single (3-2) move implements the shelling with two free faces in the boundary.

Finally, we have to deal with the situation where the 3-simplex we want to shell has three of its faces in the boundary 2-sphere. The top 3-simplex on the left of Figure 12 is the one we want to shell next, while the other three are contained in the cone. It is obvious that a single (4-1) Pachner move does the job.

Putting all these facts together, we’ve seen that in the setup described above, each elementary shelling corresponds to a single Pachner move. So if we want to bound the number of Pachner moves required for the simplification of the subdivision $S$, all we need to do is to count the number of tetrahedra in $S$. This will be dealt with in Section 6.
Before we go on to discuss the two dimensional case, we need to prove the following slightly technical lemma which connects collapsing of an edge with Pachner moves. It will be of use to us in Section 5.

**Lemma 4.1.** Let \( x \) be a vertex in a combinatorial triangulation of \( S^2 \) containing \( n \) 2-simplices. Assume further that the star of \( x \) is an embedded PL disc, triangulated by \( k \) triangles. Let \( e \) be the unique edge in the 3-ball, triangulated as a cone on \( S^2 \), running between \( x \) and the cone point. The triangulation of the same 3-ball obtained by crushing the edge \( e \), and thus flattening its star, can be constructed by \( (n-k+1) \) Pachner moves used on the original (coned) triangulation.

**Proof.** The 3-ball from the lemma can be view as a union of the following two PL 3-balls: The star of the edge \( e \) and the cone on the disc in the bounding \( S^2 \), which is the complement of the star of the vertex \( x \) on the 2-sphere.

The triangulation we are aiming for is equal to the triangulation of the latter 3-ball. We therefore want to flatten the star of the edge \( e \) down to the cone on the link of \( e \).

This can be achieved by “moving” the cone points of the 3-simplices in the second of the two 3-balls described above, from our initial cone point to the vertex \( x \). Such a 3-simplex, having a face in \( S^2 \), which is adjacent to the star of \( x \), can be moved by a (2-3) move or its inverse, depending on the number of edges it has in common with the star of \( x \).

Repeating this for all (but one) 3-simplices in the cone on the disc \( S^2 - \text{int}(	ext{star}(x)) \) almost does the job. All we have to do at this stage, is to use a single (4-1) move on what’s left of the two 3-balls described above.

We should also note that the sequence of (2-3) moves and their inverses, we used to alter the initial triangulation, can always be found. This follows from the well-known fact that every combinatorial triangulation of a PL 2-disc is shellable. □
The rest of this section will be devoted to two dimensional Pachner moves and their relationship with elementary shellings. In fact, what we want to do is to transform any given triangulation of a disc into a cone on its boundary, using Pachner moves only.

In dimension two there are three possible moves at our disposal. They are given by Figure 13.

![Figure 13. Two dimensional Pachner moves.](image)

The simplifying procedure for any PL disc is described by the next lemma.

**Lemma 4.2.** Any combinatorial triangulation of a piecewise linear disc with \( n \) triangles can be altered into a cone on the boundary of the same disc by \( n \) Pachner moves.

There are two reasons why we can (and have to) assume that the triangulation of the disc is combinatorial. The first one is that in what follows, we can easily guarantee this property for all the discs we are going to be using our Lemma 4.2 on. The second one is that the proof of the above lemma relies on the fact that any triangulation of a disc is shellable, a fact not entirely correct (with our definition of an elementary shelling) if we allow for non-combinatorial triangulations.

**Proof.** Since the triangulation of our disc is shellable, we can index all the simplices in it by numbers from 1 to \( n \), so that the increasing integers specify a way of reducing our triangulation down to a single triangle. The 2-simplex that’s left has index \( n \). Let’s make a (1-3) move on it. The 2-simplex corresponding to \( n - 1 \) has to share a unique edge with it. Making a (2-2) move over this edge, changes our original triangulation in the last two 2-simplices to a cone on the boundary of the disc that they compose. The rest of the triangulation is unchanged at this stage.

Noticing that the union of the last \( k \) 2-simplices in our sequence always gives a disc, makes the following induction possible. Say that we already have a cone on the boundary of the disc which is the union of the last \( k \) 2-simplices and that the rest of the triangulation we started with is unchanged. If the triangle corresponding to \( n - k - 1 \) has a single edge in common with our cone, we act as before (a single (2-2) move suffices). If it has two faces in common, a single (3-1) move finishes the proof. \( \square \)
5. The subdivided triangulation.

Let $T$ be a possibly non-combinatorial triangulation of $S^3$ with $t$ tetrahedra. Let’s also make the following technical assumption on $T$: Each edge in the 1-skeleton of $T$ appears at most once as an edge of any 3-simplex in $T$. This assumption does not imply that the triangulation $T$ is combinatorial, but it is certainly satisfied by all combinatorial triangulations of $S^3$. We are making it at this stage because it is going to simplify some of the processes we’ll have to invoke later on. It will also become clear that any triangulation can be altered so that it has this property by linearly (in $t$) many Pachner moves.

In this section, we shall describe the subdivision $S$ of the triangulation $T$ and also bound the number of Pachner moves required to construct it.

Let $\Gamma$ be the union of all discs needed to perform all elementary and natural isotopies in all the pieces of $S^3 - \Sigma$. We should note that the number of elements of $\Gamma$, coming from a single 2-sphere in $\Lambda$, is bounded above by the number of times the almost normal sphere in question intersects the 1-skeleton. An explicit bound on the number of elements of $\Gamma$ will be given later.

An elementary disc from $\Gamma$ will intersect every tetrahedron in the star of the edge we are isotoping over in a disc region (see Figure 14).

![Figure 14](image)

**Figure 14.** Regions, in a disc in $\Gamma$, correspond to tetrahedra in the star of an edge.

In each tetrahedron, the operation of adding in this disc will consist of gluing in a length four (respectively two) disc so that two (respectively one) arcs in its boundary are contained in the isotoped almost normal surface, and the other two (respectively one) lie in the boundary of the tetrahedron.

We are now in the position to describe the subdivision of the polyhedron

$$T^2 \cup \Sigma \cup \Lambda \cup \Gamma$$

which will be a subcomplex of the triangulation $S$. In fact, the simplicial structure of the polyhedron $T^2 \cup \Sigma \cup \Lambda \cup \Gamma$ will play a crucial role in the simplifying process and will also be of significance in the definition of the subdivision $S$.

All the normal 2-spheres in $\Sigma$ will inherit the PL structure from their normal structure. The normal triangles in $\Sigma$ will become 2-simplices, while the normal quadrilaterals will be subdivided into two 2-simplices by a diagonal.
The PL structure of the almost normal 2-spheres in $\Lambda$ will be a subdivision of the normal and almost normal pieces. We will subdivide them according to the markings on them, made by discs in $\Gamma$, where we define a marking on a normal or an almost normal piece to be an arc of intersection of the piece with a disc in $\Gamma$. We know, that each element of $\Gamma$ is chopped up into discs of lengths two or four in each tetrahedron (as in Figure 14).

The disc regions of the elements of $\Gamma$ of length four can only leave a marking on a normal piece of an almost normal 2-sphere going from one normal arc to another. There can only be three such markings on a triangle and four of them on a quadrilateral, one for each corner. On an almost normal octagon, immediately after we glue on our first disc corresponding to the natural isotopy, we end up with two triangles. So there can be at most six markings on an octagon, coming from the discs of length four.

The discs of length two will either leave a marking running from a normal arc to some other marking or simply running between two markings. Because each marking is parallel to some edge of the normal piece that it lies on and because we can not get more than one marking of the same kind, superimposing all the possible markings on normal and almost normal pieces is described by Figure 15.

![Figure 15. The polyhedral structure of normal and almost normal pieces of elements in $\Lambda$.](image)

The almost normal piece which is obtained by tubing together two normal pieces can be treated in the same way, since we could view it as an annulus around an edge between two normal 2-spheres. This annulus consists of discs of length four in each tetrahedron in the star of the connecting edge. These discs will be glued on the pairs of normal pieces yielding non-normal pieces, similar to the ones we get during the isotopy of the surface $F$ from Lemma 2.7. The PL structure on such a piece will come from the PL structure on the two parts of normal pieces it consists of, and from the
PL structure on the glued in disc, which we haven’t yet described. These glued in discs from the annulus behave in the same way as the disc regions from elements in \( \Gamma \) (see Figure 14).

In the next paragraph we shall see that each of these disc regions can be triangulated by at most 6 triangles. Counting the regions in the normal and almost normal pieces in Figure 15 and triangulating each region (if it is not a triangle already) by coning from one of the vertices in its boundary, we can see that each piece, including the ones coming from the “tubed” almost normal 2-spheres, contains less than 200 2-simplices. We should also note that the described subdivision of the pieces is combinatorial.

Now, we have to put a PL structure on the elements of \( \Gamma \). We’ve noted before (Figure 14) that each elementary disc in \( \Gamma \) consists of disc regions of lengths two or four. Once we’ve glued in a disc from \( \Gamma \), the disc regions in it give us a polyhedral structure on it. Further gluings will however subdivide this structure. Concentrating on a single disc region \( A \) of our element in \( \Gamma \), we note that all further gluings of disc regions of length four will miss \( A \) completely and therefore not change it at all. Disc regions of length two can add in a further arc on \( A \) which runs parallel to the arc(s) in its boundary, contained in the 2-skeleton \( T^2 \). Since this can only happen once per boundary arc of \( A \) in the 2-skeleton \( T^2 \), we can add at most two arcs in each disc region of any element in \( \Gamma \). So a disc in \( \Gamma \) will in the end look exactly like the disc in Figure 14 with less than 3\( t \) disc regions. This follows from the assumption we made at the beginning of this section, since it implies that a star of an edge can contain at most \( t \) tetrahedra.

The arcs in the boundaries of disc regions of elements in \( \Gamma \) that leave markings on normal and almost normal pieces of the elements in \( \Lambda \) will be subdivided further by the vertices coming from the points of intersection of the markings (see Figure 15). An arc in the boundary of the length two disc region (i.e., the one that’s leftmost or rightmost in Figure 14) will get at most 16 vertices in this way, while an arc in the boundary of the length four disc region will contain at most 5 such vertices (see Figure 15).

All these observations about the polyhedral structure of the discs in \( \Gamma \) imply that each disc region corresponding to a single tetrahedron in \( T \), will be triangulated by no more than 20 triangles. So we can triangulate any element from \( \Gamma \) by less than 20\( t \) triangles. Again, the triangulation we get is combinatorial.

Finally, we need to induce a PL structure on the 2-skeleton \( T^2 \). Normal and almost normal simple closed curves bounding pieces of elements of \( \Sigma \) and \( \Lambda \) will partition the 2-skeleton \( T^2 \) into piecewise linear regions and thus induce a polyhedral structure on it. We only have five nontrivial complementary regions in the boundary of every tetrahedron in \( T \). They are as in Figure 16.
Figure 16. Regions in the boundary of a 3-simplex bounded by normal and almost normal simple closed curves.

So topologically we have two annuli, two twice punctured discs and one three times punctured disc. The technical assumption on the triangulation $T$, we made at the beginning of this section, implies that all these surfaces are embedded in the 3-sphere.

We also need to take into account the discs in $\Gamma$ which will subdivide further the polyhedral structure that the surfaces in Figure 16 already have. Each disc region of an element in $\Gamma$ will give a further arc in one of the regions in Figure 16. This arc will run from one normal arc in the boundary of the region to the other. Its end points are vertices of the subdivision of normal and almost normal pieces of the 2-sphere elements from $\Lambda$. It is worth noting that an arc in the boundary of a tetrahedron, coming from a disc in the family $\Gamma$, will neither connect two segments in the 1-skeleton $T^1$ nor will it connect a normal arc with a segment in $T^1$. Since a normal arc can have at most 4 vertices in its interior (Figure 15), it follows that we will never have to add in more than 50 arcs per planar surface (adding up all the possibilities in all the regions in Figure 16) in $T^2$.

We can now obtain the simplicial structure on the 2-skeleton just by coning from one of the vertices of each disc subregion of the planar surfaces in Figure 16. It now follows that each surface in Figure 16 is triangulated by less than 200 2-simplices.

Now we are in the position to describe completely the subdivision $S$ of the triangulation $T$ we started with. As it was said before, the polyhedron $T^2 \cup \Sigma \cup \Lambda \cup \Gamma$ with its simplicial structure is going to be a subcomplex of $S$. Lemma 2.7 tells us that the complement of the polyhedron $T^2 \cup \Sigma \cup \Lambda \cup \Gamma$ in each tetrahedron of $T$ is just a union of 3-balls. The boundary 2-spheres of these 3-balls are embedded by the assumption we made at the very beginning of this section. They also inherit a PL structure from $T^2 \cup \Sigma \cup \Lambda \cup \Gamma$ which is combinatorial. Every complementary 3-ball can thus be triangulated by adding a vertex in its interior and coning its boundary. Since these 3-balls
exhaust the whole 3-sphere, the cones completely determine the subdivision $S$. We should also note that all these coned 3-balls are in fact shellable because their bounding 2-spheres are triangulated in a combinatorial fashion.

The rest of this section will be devoted to obtaining the subdivision $S$ from the triangulation $T$ using Pachner moves. The basic tool for achieving this end will be the procedure called changing of cones.

Suppose we had two PL discs $D$ and $E$ with isomorphic simplicial structure on their boundaries. Let the union $D \cup E$ denote the PL 2-sphere obtained by gluing the two discs together via a simplicial isomorphism on their boundaries. What we want is an algorithm to transform the cone on $D$, denoted by $CD$, to the union of cones $CE \cup C(D \cup E)$, without changing the triangulation of $D$. This is described schematically by Figure 17.

![Figure 17. The changing of cones.](image)

We have the following lemma giving a bound on the number of Pachner moves required for changing of cones.

**Lemma 5.1.** Let discs $D$ and $E$ be as above, where $n$ is the number of 2-simplices in $D$ and $m$ is the number of 2-simplices in $E$. Then we can perform the changing of cones using less than $4(n + m)$ Pachner moves.

**Proof.** We will divide the process into three steps. First, we glue a cone on the cone on the boundary of $D$ onto the bottom part of the boundary of $CD$ (Figure 18). This is a reversed process to destroying an edge which connects the two cone points of the bit that we glued on. It can therefore, by Lemma 4.1, be accomplished by less than $(n + 1)$ Pachner moves.

In the second step we perform the same move again, i.e., we glue the cone $C(C(\partial D))$ onto the space we’ve got so far (Figure 18). This again requires not more than $(n + 1)$ Pachner moves.

The space we’ve created can be described as a suspension of $C(\partial D)$ glued onto the cone on the disc $D$. We know that we can transform the cone triangulation of the disc $C(\partial D)$ into the triangulation of $E$ by using not more than $(n + m)$ two dimensional Pachner moves (Lemma 4.2). It is also
clear that in the suspension setting, each (1-3) move (or its inverse) can be realized by one (1-4) and one (2-3) move. A (2-2) Pachner move can be realized by a (2-3) and a (3-2) move. Putting all this together implies our bound.

□

The changing of cones will help us produce all the necessary cones in the triangulation $S$. Now we have at our disposal all the tools required, to bound the number of Pachner moves needed for obtaining the subdivision $S$ from the triangulation $T$.

The whole process will be divided into five stages. We’ll start by describing each one of them, and then we’ll bound the number of moves we made.

1) Add a vertex into every tetrahedron and every triangle of the triangulation $T$ and cone.
2) Subdivide the 1-skeleton of $T$ to get a subcomplex of $S$, and keep the triangulation in the 3-simplices of $T$ coned.
3) Subdivide the 2-skeleton of $T$ to get a subcomplex of $S$, and keep the triangulation in the 3-simplices of $T$ coned.
4) Chop up tetrahedra of $T$ by the appropriate normal and almost normal pieces and triangulate the complementary regions by coning them from a point in their interior.
5) Chop up the complementary regions of 4 by length two and length four disc regions of elements in $Γ$. Cone the complements.

We note that Step 3 can be accomplished by suspending the process in Lemma 4.2. Steps 4 and 5 are possible by Lemma 5.1.

Adding a vertex into each 3-simplex in $T$ takes $t(1-4)$ moves. Adding one into a triangle of $T$ takes two Pachner moves: One (1-4) move followed by a (2-3) move. So Step 1 amounts to $5t$ Pachner moves since there are precisely $2t$ triangles in the triangulation $T$.

We should note that the subdivision we get after Step 1 will always satisfy the technical condition we stipulated at the beginning of this section. This
is simply because every tetrahedron of this subdivision contains precisely one edge from the 1-skeleton of $T$. Its other edges are embedded in the 2-simplices and in the tetrahedra of $T$. It is also clear that this subdivision contains $12t$ 3-simplices. So the worst case scenario would make us do Step 1 at the very beginning and then do the simplification process (that we’ve been describing) on that subdivision. So once we work out the bound for this simplification procedure, we have to substitute each $t$ in the formula with $12t$.

Let’s go back to the construction of the subdivision $S$. First we want to bound the number of vertices of $S$ in each edge of the triangulation $T$. By Lemma 2.6 it follows that there are at most $3 \cdot 2^{110t^2}$ normal arcs in any triangle of $T$, coming from all elements in $\Sigma$ and $\Lambda$. Since each normal arc contributes at most one point of intersection with a single edge, we will have less than $3 \cdot 2^{110t^2}$ vertices on any edge in the 1-skeleton $T^1$. Since there are less than $5t$ edges all together (an Euler characteristic count), the number of vertices of the triangulation $S$, contained in $T^1$ will be bounded by $15t^2 2^{110t^2}$.

The star of any edge in $T$ contains at most $2t$ 3-simplices in the subdivision we have so far. Creating a vertex on this edge can obviously be done in the following way: First make a (1-4) move on one of the simplices in the star of the edge. Then do a sequence, of length at most $2t - 2$, of (2-3) Pachner moves. Now the addition of the vertex can be finished off by a single (3-2) Pachner move. All together this procedure takes not more than $2t$ Pachner moves. Step 2 will thus require no more than $30t^2 2^{110t^2}$ Pachner moves.

We already know that there will be at most $3 \cdot 2^{110t^2}$ normal arcs in any triangle of $T$. So the number of regions in a 2-simplex in the 2-skeleton $T^2$ is therefore bounded by the same number (plus one). These regions correspond to the regions in the surfaces from Figure 16 and will thus be triangulated by less than 20 2-simplices. So any triangle in $T$ will be subdivided by at most $60 \cdot 2^{110t^2}$ 2-simplices. By Lemma 4.2, this configuration can be obtained by $60 \cdot 2^{110t^2}$ two dimensional Pachner moves (we should notice here that before starting the process from Lemma 4.2, the triangles of $T$ were subdivided as cones on their boundaries). Suspending this process and doing it for all $2t$ 2-simplices in $T^2$ yields an upper bound of $3 \cdot 10^2 t^2 2^{110t^2}$ Pachner moves used in Step 3. This is because every two-dimensional Pachner move requires 2 three-dimensional ones.

The number of 3-ball regions, the elements of $\Sigma$ and $\Lambda$ produce in all tetrahedra of $T$, is equal to the number of normal and almost normal pieces in all the 2-spheres from $\Sigma$ and $\Lambda$ (plus $t$). So it is bounded above by
3 \cdot 2^{110t^2}. Using Lemma 5.1, we are going to change the cone structure in every tetrahedron in $T$. This will be accomplished, step by step, starting from the vertices of the tetrahedron and moving towards the cone point in its interior. At each stage we have to change a disc consisting of one of the surfaces in Figure 16, where all but one of its boundary components already have their corresponding normal and almost normal pieces glued in (that makes it a disc), to a disc coming from the only normal or almost normal piece that hasn’t yet been introduced. Since we want the region between the two discs we’ve just described, to be coned, Lemma 5.1 is precisely what is needed. It is also obvious that the disc $D$ from Lemma 5.1 will in this situation never contain more than 800 triangles (this follows from the counts we did when defining the subdivision $S$), while the disc $E$, which is just a normal or an almost normal piece, will be triangulated by less than 200 2-simplices. So in a single 3-ball region, we'll make less than $4 \cdot (800 + 200)$ Pachner moves (Lemma 5.1). In order to complete Step 4 in all the tetrahedra of $T$, we need to make

$$12 \cdot 10^3 2^{110t^2}$$

Pachner moves.

The number of discs in $\Gamma$, coming from a single element in $\Lambda$, is bounded above by half the number of times the 2-sphere in question intersects the 1-skeleton. We already know that there are at most $3 \cdot 2^{110t^2}$ vertices on any edge in the 1-skeleton of the triangulation $T$. Since there are less then $5t$ edges in $T^1$, the number of elements in $\Gamma$ is bounded above by $\frac{1}{2} 15t 2^{110t^2} < 10t 2^{110t^2}$.

Each of the discs in $\Gamma$ has at most $t$ disc regions (by the assumption from the beginning of this section), coming from the 3-simplices in the star of the edge the particular disc corresponds to. Each disc region is triangulated by strictly less than 20 triangles. A disc region in an element of $\Gamma$ will correspond to the disc $E$ in Lemma 5.1.

The boundary of each disc region is a subcomplex in the boundary of a coned 3-ball. One of the complementary discs bounded by this simple closed curve, in the boundary of the coned 3-ball, will correspond to the disc $D$ in Lemma 5.1. In the case of a disc region in an element of $\Gamma$ having two arcs in its boundary embedded in the 2-skeleton $T^2$, the disc corresponding to $D$ we were discussing before will contain six 2-simplices (two in normal or almost normal pieces and four in the 2-skeleton $T^2$).

Let’s look at the case of a disc region from an element in $\Gamma$ that intersects the 2-skeleton of $T$ in a single arc (i.e., the leftmost or the rightmost region in Figure 14) and corresponds to an elementary isotopy. The number of triangles of the complementary region (in the bounding 2-sphere) we are interested in will then be smaller than the sum of the numbers of 2-simplices in the following surfaces: The disc in the 2-simplex of $T$ our disc region is
parallel to, the disc in the 2-simplex of $T$ containing a bounding arc of the
disc region we are gluing in, regions in at most three normal triangles or
regions in a normal triangle and a normal quadrilateral or regions in two
normal quadrilaterals, at most two discs contained in two distinct regions
in the elements of $\Gamma$. Bounds for the numbers of 2-simplices for the above
surfaces are as follows: 20, 2, 3 · 30 or 2 · 30 or 2 · 30, 2 · 2 respectively. What
happens with the disc regions belonging to the elements of $\Gamma$ that come from
natural isotopies? In that case the disc $D$ from Lemma 5.1 is composed of
the following surfaces: Roughly a half of an almost normal octagon, three
discs contained in the 2-simplices of $T$, a single normal triangle. The explicit
bounds in this case are: 70, 3 · 20, 70.

An upper bound on the sum of the numbers of triangles in $D$ and $E$
will therefore always be strictly less than 300 (we already know that a disc
region in an element from $\Gamma$ contains no more then 20 2-simplices). So by
Lemma 5.1, we can produce our disc region in this 3-ball by less than 4 · 300
Pachner moves. All together, we have to make less than

$$12 \cdot 10^3 t^2 2^{110^2}$$

Pachner moves in order to complete Step 5.

Summing everything up, estimating the resulting expression and substituting $t$ with 12$t$ to account for the technical assumption we made at the
beginning of this section, we get the following proposition.

**Proposition 5.2.** Let $T$ be any triangulation of the 3-sphere and let $t$ be
the number of tetrahedra in it. Then the subdivision $S$, described at the
beginning of this section, can be obtained from $T$ by making less than $ct^2 2^{dt^2}$
Pachner moves, where the constant $c$ is bounded above by $5 \cdot 10^6$ and the
constant $d$ is smaller than $2 \cdot 10^4$.

### 6. Conclusion of the proof.

Now, we are in the position to bound the number of Pachner moves needed
to simplify any given triangulation $T$ of the 3-sphere, down to the canonical
triangulation with only two tetrahedra. We will apply the shelling tech-
niques, developed in Section 4, to the subdivision $S$ of the triangulation $T$,
described in Section 5.

The basic question we have to answer at this point is how many tetrahedra
do we have to shell in the simplifying process. Then we can estimate the
number of Pachner moves needed for the process, using the fact that each
elementary shelling corresponds to a single Pachner move.

Let’s bound first the total number of tetrahedra of $S$.

This will be accomplished in two steps. First we count the number of
3-ball regions we coned, while constructing the subdivision $S$, in all the
tetrahedra of the triangulation $T$. The second step consists of bounding
the number of triangles in each of the boundaries of the 3-balls mentioned above. Multiplying these two numbers gives our bound.

Lemma 2.6 implies that there are at most $3 \cdot 2^{110t^2}$ normal and almost normal pieces in all 3-simplices of $T$, coming from all normal and almost normal 2-sphere in $\Sigma \cup \Lambda$. We know that each piece contains at most 200 triangles. Each planar surface in the boundary of the tetrahedron (see Figure 16) contains at most 50 arcs and is triangulated by at most 200 triangles. Each 3-ball component of the complement of $\Sigma \cup \Lambda$ in our tetrahedron will thus contain less than 50 disc regions, coming from elements in $\Gamma$.

So in all 3-simplices of $T$ we’ll have not more than $50 \cdot 3 \cdot 2^{110t^2}$ 3-ball regions. Since each disc region in any element in $\Gamma$ contains less than 20 triangles, 1000 is surely an upper bound on the number of triangles in the boundary of any of the 3-ball regions. There will therefore be at most $15 \cdot 10^4 2^{110t^2}$ tetrahedra in $S$.

Combining Proposition 5.2 and the assumption that there are precisely $12t$ tetrahedra in the triangulation $T$, concludes the proof of the main theorem.

Acknowledgements. I would like to thank my research supervisor Marc Lackenby for many helpful (and enjoyable) conversations about math and about everything else.

References

Received January 30, 2001 and revised October 27, 2001.

Department of Pure Mathematics and Mathematical Statistics
Center for Mathematical Sciences
University of Cambridge
Wilberforce Road
Cambridge, CB3 0WB
UK
E-mail address: a.mijatovic@dpmms.cam.ac.uk
PAYNE–POLYÀ–WEINBERGER TYPE INEQUALITIES FOR EIGENVALUES OF NONELLIPTIC OPERATORS

Pengcheng Niu and Huiqing Zhang

In this paper we consider the eigenvalue problems for some nonelliptic operators which include the real Kohn-Laplacian in the Heisenberg and generalized Baouendi-Grushin operator. Some interesting inequalities for eigenvalues are given by establishing the identities and inequalities for noncommutative vector fields.

1. Introduction.

Let $\Delta$ denote the Laplacian in the Euclidean space. The classic upper estimates, independent of the domain, for the gaps of eigenvalues of $-\Delta$, $(-\Delta)^2$ and $(-\Delta)^k (k \geq 3)$ were studied extensively by many mathematicians, cf. Payne, Polya and Weinberger [16], Hile and Yeh [10], Chen and Qian [2], Guo [8] etc.. The asymptotic behaviors of eigenvalues for degenerate elliptic operators were considered by Beals, Greiner and Stanton [1], Menikoff and Sjöstrand [15], Fefferman and Phane [3, 4], Garofalo and Shen [7], respectively.

In this paper, we are concerned with the following eigenvalue problem

\begin{align}
-\Delta_{H_n} u &= \lambda u, \quad \text{in } \Omega, \quad u = 0, \quad \text{on } \partial \Omega \tag{1.1} \\
(-\Delta_{H_n})^2 u &= \lambda u, \quad \text{in } \Omega, \quad u = \frac{\partial u}{\partial \nu} = 0, \quad \text{on } \partial \Omega \tag{1.2} \\
(-\Delta_{H_n})^k u &= \lambda u, \quad \text{in } \Omega, \quad u = \frac{\partial u}{\partial \nu} = \cdots = \frac{\partial^{k-1} u}{\partial \nu^{k-1}} = 0, \quad \text{on } \partial \Omega \tag{1.3}
\end{align}

where $\Omega$ is a bounded domain in the Heisenberg group $H_n$, with smooth boundary $\partial \Omega$, $n \geq 1, \nu$ is the unit outward normal to $\partial \Omega$, $k \geq 3$ is a positive integer. Let $\Delta_{H_n}$ denote the real Kohn-Laplacian in the Heisenberg group $\sum_{j=1}^n (X_j^2 + Y_j^2)$, where $X_j = \frac{\partial}{\partial x_j} + \frac{y_j}{2} \frac{\partial}{\partial \eta}, Y_j = \frac{\partial}{\partial y_j} - \frac{x_j}{2} \frac{\partial}{\partial \eta}, j = 1, \ldots, n, T = \frac{\partial}{\partial \tau}$. The existence of eigenvalue for (1.1) has been proved by Luo and Niu [12, 13, 14] using the Kohn inequality (see [11]) for the vector fields \{X_j, Y_j\} together with the spectral properties of compact operators. In what follows we let

\begin{equation}
0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m \leq \cdots \to +\infty \tag{1.4}
\end{equation}

325
denote the successive eigenvalues for (1.1) with corresponding eigenfunctions \(u_1, u_2, \ldots, u_m, \ldots\) in \(S^{1,2}_0(\Omega)\). Here, \(S^{1,2}(\Omega)\) denotes the Hilbert space of the functions \(u \in L^2(\Omega)\) such that \(X_j u, Y_j u \in L^2(\Omega)\), and \(S^{1,2}_0(\Omega)\) denotes closure of \(C_0^\infty(\Omega)\) in the norm

\[
\|u\|_{S^{1,2}}^2 = \int_\Omega (|\nabla_{H_n} u|^2 + |u|^2) dx dy dt
\]

where \(\nabla_{H_n} u = (x_1 u, \ldots, X_n u, Y - 1 u, \ldots, Y_n u)\) normalized so that

\[
\langle u_i, u_j \rangle = \int_\Omega u_i u_j dx dy dt = \delta_{ij}, \quad i, j = 1, 2, \ldots.
\]

For simplicity we will leave out \(\Omega\) and \(dxdydt\) in all integrals in the sequel and denote \(L = -\triangle_{H_n}\).

It is clear that these statements are also valid for the problems (1.2) and (1.3) and the eigenfunctions belong to \(S^{2,2}_0(\Omega)\) and \(S^k_0(\Omega)\) respectively.

We will derive some upper estimates which are independent of the domain, for the eigenvalues of (1.1), (1.2), and (1.3), respectively. The noncommutativity of vector fields \(\{X_j, Y_j\}\) makes the discussion of these problems more complicated than one in the case of Euclidean-Laplacian. Furthermore, we will also consider the eigenvalues of generalized Baouendi-Grushin operators [6].

The paper is constructed as follows: Section 2 presents some identities and inequalities based on the vector fields \(\{x_j, Y_j, T\}\), \(j = 1, \ldots, n\), which show the reason that the problems (1.1), (1.2) and (1.3) are treated separately. The main estimates for the eigenvalues of (1.1), (1.2) and (1.3) are given in Section 3, Section 4 and Section 5 respectively. We conclude Section 6 by estimating eigenvalues of the generalized Baouendi-Grushin operator.

2. Some preliminary lemmas.

We establish some properties for the commutative vector fields \(\{x_j, Y_j, T\}\), \((j = 1, \ldots, n)\) which are of independent interest.

**Lemma 2.1.** Given any positive integer \(p, 1 \leq p \leq k\), we have

\[
L^p \begin{pmatrix} x_j u_i \\ y_j u_i \end{pmatrix} = \begin{pmatrix} x_j \\ y_j \end{pmatrix} L^p u_i - 2 \begin{pmatrix} X_j \\ Y_j \end{pmatrix} L^{p-1} u_i - 2 \sum_{q=1}^{p-1} p - 1 L^{p-q} \begin{pmatrix} X_j \\ Y_j \end{pmatrix} L^{q-1} u_i
\]

\(i = 1, \ldots, m, \ldots, j = 1, \ldots, n\).

**Proof.** Since

\[
X_1(x_1 u_i) = u_i + x_1 X_1 u_i,
\]

\[
X_j(x_1 u_i) = x_1 X_j u_i, \quad j = 2, \ldots, n,
\]

\[
Y_j(x_1 u_i) = x_1 Y_j u_i, \quad j = 2, \ldots, n,
\]
and
\[
X_1^2(x_1 u_i) = 2X_1 u_i + x_1 X_1^2 u_i, \\
X_j^2(x_1 u_i) = x_1 X_j^2 u_i, \quad j = 2, \ldots, n, \\
Y_j^2(x_1 u_i) = x_1 Y_j^2 u_i, \quad j = 2, \ldots, n,
\]
we obtain
\[
(2.2) \quad L(x_1 u_i) = x_1 Lu_i - 2X_1 u_i
\]
and so
\[
(2.3) \quad L^p(x_1 u_i) = L^{p-1}(x_1 Lu_i - 2X_1 u_i) = L^{p-1}(x_1 Lu_i) - 2L^{p-1}X_1 u_i.
\]
Changing \(u_i\) in (2.2) to \(Lu_i\) yields
\[
L(x_1 Lu_i) = x_1 L^2 u_i - 2X_1 Lu_i
\]
and so
\[
L^{p-1}(x_1 Lu_i) = L^{p-2}(x_1 L^2 u_i) - 2L^{p-2}X_1 Lu_i,
\]
Substituting into (2.3) and repeating these steps, we prove the first formula in (2.1) for \(L^p(x_1 u_i)\). \(\square\)

**Remark 2.1.** When \(p = 1, 2\), we have
\[
(2.4) \quad L \begin{pmatrix} x_j u_i \\ y_j u_i \end{pmatrix} = \begin{pmatrix} x_j \\ y_j \end{pmatrix} Lu_i - 2 \begin{pmatrix} x_j \\ Y_j \end{pmatrix} u_i
\]
\[
(2.5) \quad L^2 \begin{pmatrix} x_j u_i \\ y_j u_i \end{pmatrix} = \begin{pmatrix} x_j \\ y_j \end{pmatrix} L^2 u_i - 2 \begin{pmatrix} x_j \\ Y_j \end{pmatrix} Lu_i - 2L \begin{pmatrix} x_j \\ Y_j \end{pmatrix} u_i
\]
respectively. It yields the main difference in the following estimations.

Let
\[
0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m \leq \ldots
\]
denote the eigenvalues of (1.1) (1.2) and (1.3), respectively) with corresponding orthogonal normalized eigenfunctions \(u_1, u_2, \ldots, u_m, \ldots\) in \(S_{1,2}^0(\Omega)\) \((S_{0,2}^2(\Omega)\) and \(S_{k,2}^0(\Omega)\), respectively).

Take the trial functions
\[
(2.6) \quad \varphi_{ixj} = x_j u_i - \sum_{i=1}^m a_{iixj} u_i, \quad \varphi_{iyj} = y_j u_i - \sum_{i=1}^m a_{iyj} u_i
\]
where \(a_{iixj} = \int x_j u_i u_i, a_{iyj} = \int y_j u_i u_i\). It is easy to know that each function of (2.6) is orthogonal to \(u_1, \ldots, u_m\) and vanishing on \(\partial \Omega\).

**Lemma 2.2.** For \(m \geq 1\), it holds
\[
(2.7) \quad \sum_{i=1}^m \int \varphi_{ixj} X_j u_i = \sum_{i=1}^m \int \varphi_{iyj} Y_j u_i = \frac{-m}{2}, \quad j = 1, \ldots, n.
\]
Proof. We only prove the first equality in (2.7). Since \( a_{ilx_j} = a_{ilx_j} \), \( \int_{\Omega} u_i x_j u_i \)

\(-\int_{\Omega} u_i X_j u_i \) and thus \( \sum_{i,l} a_{ilx_j} \int u_i X_j u_i = 0 \), one has

\[
\sum_{i} \int \varphi_{ix_j} X_j u_i = \sum_{i} \int (x_i u_i - \sum_{i} a_{ilx_j} u_i) X_j u_i
\]

\[
= \sum_{i} \int x_i u_i X_j u_i - \sum_{i} \int X_j (x_i u_i) u_i
\]

\[
= - \sum_{i} \int u_i^2 - \sum_{i} \int x_i u_i X_j u_i
\]

\[
= -m - \sum_{i} \int x_i u_i X_j u_i.
\]

The desired equality is derived immediately. \( \square \)

For simplicity, we let \( u_i \in C_0^\infty(\Omega) \) in what follows. Obviously, by the density property, the following results are valid for \( u_i \in S_0^{1,2}(\Omega) \).

**Lemma 2.3.** For any \( p \geq 1 \), we have

\[
\left( \int_{\Omega} |\nabla^p L u_i|^2 \right)^{\frac{1}{p}} \leq \left( \int_{\Omega} |\nabla^{p+1} L u_i|^2 \right)^{\frac{1}{p+1}}, i = 1, \ldots, m, \ldots
\]

where \( \nabla L u = (X_1 u, \ldots, X_n u, Y_1 u, \ldots, Y_n u) \).

Proof. Evidently, for \( p \geq 1 \)

\[
\int |\nabla L u_i|^2 = - \int u_i L u_i \leq \left( \int u_i^2 \right)^{\frac{1}{2}} \left( \int (L u_i)^2 \right)^{\frac{1}{2}} = \left( \int |\nabla^2 L u_i|^2 \right)^{\frac{1}{2}}
\]

by the induction assumption it follows that

\[
\int |\nabla^p L u_i|^2 = - \int \nabla^{p-1} L u_i \nabla^{p+1} L u_i
\]

\[
\leq \left( \int |\nabla^{p-1} L u_i|^2 \right)^{\frac{1}{2}} \left( \int |\nabla^{p+1} L u_i|^2 \right)^{\frac{1}{2}}
\]

\[
\leq \left( \int |\nabla^p L u_i|^2 \right)^{\frac{p-1}{2p}} \left( \int |\nabla^{p+1} L u_i|^2 \right)^{\frac{1}{2}}.
\]

Therefore the assertion of the Lemma is proved. \( \square \)

As a consequence, we have:

**Corollary 2.1.** For \( k \) in (1.3), we have

\[
\int |\nabla L u_i|^2 \leq \left( \int |\nabla^k L u_i|^2 \right)^{\frac{1}{k}}.
\]
Lemma 2.4. For $p \geq 1$,

\[
L^p \left( \frac{X_j}{Y_j} \right) = \sum_{s=0}^{p} C_p^s \left( (-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} A_j^s \right) L^{p-s}(2T)^s,
\]

where $j = 1, \ldots, n$; $[\bullet]$ denote the largest integer part of $\bullet$, $C_p^s = \frac{p!}{s!(p-s)!}$,

$A_j^s = X_j$, if $s = 0, 2, 4, \ldots$; $A_j^s = Y_j$, if $s = 1, 3, \ldots$.

Proof. A direct calculation gives

\[
LX_j = X_j L - 2TY_j, \quad LY_j = Y_j L - 2TX_j, \quad j = 1, \ldots, n.
\]

By the induction assumption that

\[
L^{p-1} X_j = \sum_{s=0}^{p-1} p - 1 C_{p-1} s (-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} A_j^s L^{p-s-1}(2T)^s,
\]

it follows

\[
L^p X_j = L \left[ \sum_{s=0}^{p-1} C_{p-1} s (-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} A_j^s L^{p-s-1}(2T)^s \right]
\]

\[
= X_j L^P - 2TY_j L^{p-1}
\]

\[
+ \sum_{s=2 (s \text{ even})}^{p-1} C_{p-1} s (-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} (X_j L - 2TY_j) L^{p-s-1}(2T)^s
\]

\[
+ \sum_{s=1 (s \text{ odd})}^{p-1} C_{p-1} s (-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} (Y_j L - 2TX_j) L^{p-s-1}(2T)^s
\]

\[
= X_j L^P - 2TY_j L^{p-1}
\]

\[
+ \sum_{s=2 (s \text{ even})}^{p-1} C_{p-1} s (-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} X_j L^{p-s}(2T)^s
\]

\[- \sum_{s=2 (s \text{ even})}^{p-1} C_{p-1} s (-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} Y_j L^{p-s-1}(2T)^{s+1}
\]

\[
+ \sum_{s=1 (s \text{ odd})}^{p-1} C_{p-1} s (-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} X_j L^{p-s}(2T)^s
\]

\[- \sum_{s=1 (s \text{ odd})}^{p-1} C_{p-1} s (-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} Y_j L^{p-s-1}(2T)^{s+1}
\]
\[ \begin{align*}
&= \sum_{s=0(s \text{ even})}^{p-1} \left[ C_{p-1}s(-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} X_j L^{p-s}(2T)^s 
+ C_{p-1}^{s-1}(-1)^{\left\lfloor \frac{s}{2} \right\rfloor} Y_j L^{p-s}(2T)^s \right] \\
&+ \sum_{s=1(s \text{ odd})}^{p-1} \left[ -C_{p-1}s - (-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} Y_j L^{p-s}(2T)^s 
+ C_{p-1}^{s}(-1)^{\left\lfloor \frac{s+1}{2} \right\rfloor} X_j L^{p-s}(2T)^s \right] \\
&= \sum_{s=0}^{p} C_{p}^{s}(-1)^{\left\lfloor \frac{s}{2} \right\rfloor} A_j L^{p-s}(2T)^s,
\end{align*} \]

where we have used that \( C_{p-1}s + C_{p-1}s - 1 = C_{p}^{s} \) in the last equality; \( \left\lfloor \frac{s+1}{2} \right\rfloor = \left\lfloor \frac{s}{2} \right\rfloor \) if \( s \) is even and \( \left\lfloor \frac{s+1}{2} \right\rfloor = \left\lfloor \frac{s}{2} \right\rfloor + 1 \) if \( s \) is odd. This proves the first equality in (2.10). Another equality is proved similarly. \( \square \)

**Lemma 2.5.** For \( p \geq 1 \), it has

\[ \int TL^{p+1}u_i \cdot TL^{p}u_i \leq \frac{1}{2n-1} \int L^{p+2}u_i \cdot L^{p+1}u_i, \tag{2.11} \]

\[ \int TL^{p}u_i \cdot TL^{p}u_i \leq \frac{1}{2(2n-1)} \left[ \int L^{p+2}u_i \cdot L^{p+1}u_i + \int L^{p+1}u_i \cdot L^{p}u_i \right]. \tag{2.12} \]

**Proof.** Since \( tu_i = Y_j X_j u_i - X_j Y_j u_i \), \( j = 1, \ldots, n \), we get

\[ 2n \int TL^{p+1}u_i \cdot TL^{p}u_i \]

\[ = 2 \int \sum_{j=1}^{n} (Y_j X_j - X_j Y_j) L^{p+1}u_i \cdot TL^{p}u_i \]

\[ = 2 \sum_{j=1}^{n} \int (TX_j L^{p}u_i \cdot Y_j L^{p+1}u_i - TY_j L^{p}u_i \cdot X_j L^{p+1}u_i) \]

\[ \leq \sum_{j=1}^{n} \int \left[ (TX_j L^{p}u_i)^2 + (Y_j L^{p+1}u_i)^2 + (TY_j L^{p}u_i)^2 + (X_j L^{p+1}u_i)^2 \right] \]

\[ = \sum_{j=1}^{n} \int \left[ -TX_j^2 L^{p}u_i \cdot TL^{p}u_i - Y_j^2 L^{p+1}u_i \cdot L^{p+1}u_i 
- TY_j^2 L^{p}u_i \cdot TL^{p}u_i - X_j^2 L^{p+1}u_i \cdot L^{p+1}u_i \right] \]

\[ = \int TL^{p+1}u_i \cdot TL^{p}u_i + \int L^{p+2}u_i \cdot L^{p+1}u_i. \]
and (2.11) is proved. As for (2.12), one has

\[
2 \int T L^p u_i \cdot T L^p u_i = 2 \int T \nabla L^p u_i \cdot T \nabla L^{p-1} u_i \\
\leq \int (T \nabla L^p u_i)^2 + \int (T \nabla L^{p-1} u_i)^2 \\
= \int T L^{p+1} u_i \cdot T L^p u_i + \int T L^p u_i \cdot T L^{p-1} u_i \\
\leq \frac{1}{2n-1} \left[ \int L^{p+2} u_i \cdot L^{p+1} u_i + \int L^{p+1} u_i \cdot L^p u_i \right]
\]

and the conclusion is obtained. □

**Corollary 2.2.** For positive integers \(a, p \geq 1\), the following inequalities hold:

\[\int T^a L^{p+1} u_i \cdot t^a L^p u_i \leq \frac{1}{(2n-1)^a} \int L^{p+a+1} u_i \cdot L^{p+a} u_i \] \hspace{1cm} (2.13)

\[
\int (T^a L^{p+1} u_i)^2 \leq \frac{1}{2(2n-1)^a} \left[ \int L^{p+a+1} u_i \cdot L^{p+a} u_i + \int L^{p+a} u_i \cdot L^{p+a-1} u_i \right].
\] \hspace{1cm} (2.14)

**Proof.** It is easy to obtain from (2.11)

\[
\int T^a L^{p+1} u_i \cdot t^a L^p u_i \leq \frac{1}{2n-1} \int T^{a-1} L^{p+2} u_i \cdot t^{a-1} L^{p+1} u_i \leq \ldots \\
\leq \frac{1}{(2n-1)^a} \int L^{p+a+1} u_i \cdot L^{p+a} u_i
\]

and from (2.12) and (2.13)

\[
\int (T^a L^p u_i)^2 \leq \frac{1}{2(2n-1)} \left[ \int T^{a-1} L^{p+2} u_i \cdot t^{a-1} L^{p+1} u_i \\
+ \int T^{a-1} L^{p+1} u_i \cdot t^{a-1} L^p u_i \right] \\
\leq \frac{1}{2(2n-1)^a} \left[ \int L^{p+a+1} u_i \cdot L^{p+a} u_i + \int L^{p+a} u_i \cdot L^{p+a-1} u_i \right].
\]

This completes the proof. □

### 3. Estimates of eigenvalues for (1.1).

**Theorem 3.1.** For \(m \geq 1\),

\[
\lambda_{m+1} - \lambda_m \leq \frac{2}{mn} \left( \sum_{i=1}^{m} \lambda_i \right).
\] \hspace{1cm} (3.1)
**Proof.** By the choice of trial function \( \varphi_{ix_1} \) (see (2.6)) and the Rayleigh-Ritz principle for \( \triangle H_n \), we have

\[
\lambda_{m+1} \leq \frac{\int |\nabla H_n \varphi_{ix_1}|^2}{\int |\varphi_{ix_1}|^2}, \quad i = 1, \ldots, m,
\]

and then

\[
\lambda_m \sum_{i=1}^{m} \int |\varphi_{ix_1}|^2 \leq \sum_{i=1}^{m} \int |\nabla H_n \varphi_{ix_1}|^2 = \sum_{i=1}^{m} \int L \varphi_{ix_1} \cdot \varphi_{ix_1}.
\]

By (2.4), the right-hand side becomes \(-2\int \varphi_{ix_1} X_1 u_i + \lambda_i \int \varphi_{ix_1}^2 \) and we obtain

\[
(\lambda_{m+1} - \lambda_m) \sum_{i=1}^{m} \int \varphi_{ix_1}^2 \leq -2 \sum_{i=1}^{m} \int \varphi_{ix_1} X_1 u_i.
\]

Repeating the argument to \( \varphi_{ix_j} (j = 1, \ldots, n) \) yields

\[
(\lambda_{m+1} - \lambda_m) \sum_{i=1}^{m} \sum_{j=1}^{n} \int \left( \varphi_{ix_j}^2 + \varphi_{iy_j}^2 \right) \leq -2 \sum_{i=1}^{m} \sum_{j=1}^{n} \int \left( \varphi_{ix_j} X_j u_i + \varphi_{iy_j} Y_j u_i \right).
\]

By Lemma 2.2 and Hölder’s inequality

\[
mn = - \sum_{i=1}^{m} \sum_{j=1}^{n} \int \left( \varphi_{ix_j} X_j u_i + \varphi_{iy_j} Y_j u_i \right)
\]

\[
\leq \left[ \sum_{i,j} \int \left( \varphi_{ix_j}^2 + \varphi_{iy_j}^2 \right) \right]^{\frac{1}{2}} \left[ \sum_{i,j} \int (X_j u_i)^2 + (Y_j u_i)^2 \right]^{\frac{1}{2}}
\]

\[
- \left( \sum_{i} \lambda_i \right)^{\frac{1}{2}} \left[ \sum_{i,j} \int \left( \varphi_{ix_j}^2 + \varphi_{iy_j}^2 \right) \right]^{\frac{1}{2}}
\]

hence we have

\[
\sum_{i,j} \int \left( \varphi_{ix_j}^2 + \varphi_{iy_j}^2 \right) \geq (mn)^2 \left( \sum_{i} \lambda_i \right)^{-1}.
\]

Returning to (3.2), the result is proved. \( \square \)

**Remark 3.1.** (3.1) is a generalization of Payne-Polya-Weinberger theorem for Dirichlet eigenvalues of Euclidean Laplacian to our context here.
Introducing a parameter $\alpha > \lambda_m$, we have

$$\tag{3.3} (\lambda_{m+1} - \alpha) \sum_{i,j} \int \left( \phi_{ix_j}^2 + \phi_{iy_j}^2 \right) \leq -2 \sum_{i,j} \int (\varphi_{ix_j} X_j u_i + \varphi_{iy_j} Y_j u_i)$$

$$- \sum_{i,j} (\alpha - \lambda_i) \int \left( \phi_{ix_j}^2 + \phi_{iy_j}^2 \right)$$

and for some $\delta > 0$,

$$mn = - \sum_{i,j} \int (\varphi_{ix_j} X_j u_i + \varphi_{iy_j} Y_j u_i)$$

$$\leq \frac{\delta}{2} \sum_{i,j} (\alpha - \lambda_i) \int \left( \phi_{ix_j}^2 + \phi_{iy_j}^2 \right)$$

$$+ \frac{1}{2\delta} \sum_{i,j} (\alpha - \lambda_i)^{-1} \int [(X_j u_i)^2 + (Y_j u_i)^2].$$

It is easy to see

$$\tag{4.1} (\lambda_{m+1} - \alpha) \sum_{i,j} \int \left( \phi_{ix_j}^2 + \phi_{iy_j}^2 \right) \leq 2mn - m^2 n^2 \left( \sum_{i=1}^{m} \frac{\lambda_i}{\alpha - \lambda_i} \right)^{-1}.$$ 

So an extension of Hile-Protter theorem is easily obtained:

**Theorem 3.2.** For $m \geq 1$, one has

$$\sum_{i=1}^{m} \frac{m \lambda_i}{\lambda_{m+1} - \lambda_i} \geq \frac{mn}{2}. $$

4. Estimates of eigenvalues for (1.2).

We denote the eigenvalues of (1.2) by

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m \leq \cdots \to \infty$$

with corresponding eigenfunctions $u_1, u_2, \ldots, u_m, \ldots$ in $S^2_0(\Omega)$.

**Theorem 4.1.** Let $m \geq 1$. then the following estimate holds

$$\tag{4.1} \lambda_{m+1} - \lambda_m \leq \frac{4(n + 1)}{m^2 n^2} \left( \sum_{i=1}^{m} \sqrt{\lambda_i} \right)^{1/2}.$$ 

**Proof.** Following the argument of (3.3) and noting (2.5), we have

$$\tag{4.2} (\lambda_{m+1} - \alpha)S \leq J - T,$$
where $\alpha$ is a parameter, $\alpha > \lambda_m$,
\[
S = \sum_{i,j} \left( \varphi_{ix}^2 + \varphi_{iy}^2 \right),
\]
\[
J = -2 \sum_{i,j} \int \left[ (LX_j + X_jL)u_i \cdot \varphi_{ix} + (LY_j + Y_jL)u_i \cdot \varphi_{iy} \right],
\]
\[
T = \sum_{i,j} (\alpha - \lambda_j) \int \left( \varphi_{ix}^2 + \varphi_{iy}^2 \right).
\]

We need two propositions:

**Proposition 4.1.**

\[
J \leq 4(n + 1) \sum_{i=1}^{m} \sqrt{\lambda_i}.
\]

**Proof.** Noting (2.6), one has
\[
\sum_i \int (LX_j + X_jL)u_i \cdot \varphi_{ix}
= \sum_i \int (LX_j + X_jL)u_i \left( x_ju_i - \sum_l a_{ilx_j}u_l \right)
= \sum_i \int (X_ju_iL(x_ju_i) - Lu_i \cdot X_j(x_ju_i))
- \sum_{i,l} \int (a_{ilx_j}X_ju_iLu_l - a_{ilx_j}Lu_i \cdot X_ju_i)
\triangleq M_1 + M_2.
\]

Since $a_{ilx_j} = a_{iix_j}$, $i, l = 1, \ldots, m$, we see $M_2 = 0$. On the other hand, (2.4) implies
\[
\sum_i \int (LX_j + X_jL)u_i \cdot \varphi_{ix}
= M_1 = \sum_i \int \left[ X_ju_i(-2X_ju_i + x_jLu_i) - Lu_i(u_i + x_jX_ju_i) \right]
= \sum_i \int (2X_j^2u_i \cdot u_i - u_iLu_i).
\]

Similarly, we have
\[
\sum_i \int (LY_j + Y_jL)u_i \cdot \varphi_{iy} = \sum_i \int (2Y_j^2u_i - u_iLu_i).
\]
Therefore by Hölder’s inequality we obtain
\[
J = -2 \sum_i \left[ \int -2L u_i \cdot u_i - 2n \int u_i Lu_i \right]
\]
\[
= 4(n + 1) \sum_i \int u_i Lu_i
\]
\[
\leq 4(n + 1) \sum_i \left( \int u_i^2 \right)^{\frac{1}{2}} \left( \int (Lu_i)^2 \right)^{\frac{1}{2}}
\]
\[
= 4(n + 1) \sum_i \sqrt{\lambda_i}.
\]
\[
\square
\]

**Proposition 4.2.**

\[(4.4) \quad T \geq m^2 n^2 \left( \sum_i \frac{\sqrt{\lambda_i}}{\alpha - \lambda_i} \right)^{-1}.\]

**Proof.** By Lemma 2.2 and Hölder’s inequality we have

\[(4.5) \quad mn = -\sum_{i,j} \int (\varphi_{ixj} X_j u_i + \varphi_{iyj} Y_j u_i)
\]
\[
\leq \sum_i \left[ \sum_j \int \left( \varphi_{ixj}^2 + \varphi_{iyj}^2 \right) \right]^{\frac{1}{2}} \left[ \sum_i \int \left( (X_j u_i)^2 + (Y_j u_i)^2 \right) \right]^{\frac{1}{2}}
\]
\[
\leq \frac{\delta}{2} \sum_{i,j} (\alpha - \lambda_i) \int \left( \varphi_{ixj}^2 + \varphi_{iyj}^2 \right) + \frac{1}{2\delta} \sum_i \frac{1}{\alpha - \lambda_i} \int u_i Lu_i
\]
\[
\leq \frac{\delta}{2} T + \frac{1}{2\delta} \sum_i \frac{\sqrt{\lambda_i}}{\alpha - \lambda_i},
\]

where \(\delta\) is some positive parameter. After minimizing the right-hand side of (4.5), the result is proved. \(\square\)

**Proof of Theorem 4.1.** Substituting (4.3) and (4.4) into (4.2) yields

\[(\lambda_{m+1} - \alpha)S \leq 4(n + 1) \sum_i \sqrt{\lambda_i} - m^2 n^2 \left( \sum_i \frac{\sqrt{\lambda_i}}{\alpha - \lambda_i} \right)^{-1}.\]

The inequality (4.1) is obtained with the similar discussion in [10]. \(\square\)
5. Estimates of eigenvalues for (1.3).

**Theorem 5.1.** If $k \geq 3$ is odd, then

\[
\lambda_{m+1} - \lambda_m \leq \sum_{i=1}^{m} \frac{\lambda_i^{1/2}}{m^2 n^2} \left[ (2n + 4) k \sum_{i=1}^{m} \lambda_i^{k-1} + C_1(n, k) \sum_{i=1}^{m} \lambda_i^{k-1} \right]
\]

and if $k \geq 4$ is even, then

\[
\lambda_{m+1} - \lambda_m \leq \sum_{i=1}^{m} \frac{\lambda_i^{1/2}}{m^2 n^2} \left[ (2n + 4) k \sum_{i=1}^{m} \lambda_i^{k-1} + C_2(n, k) \sum_{i=1}^{m} \lambda_i^{k-1} \right],
\]

where $C_1(n, k)$ and $C_2(n, k)$ are the constants depending on $n$ and $k$.

**Proof.** Using the trial function $\varphi_{ix_1}$ (see (2.6)) and the Rayleigh-Ritz inequality, we have

\[
\lambda_{m+1} \leq \int \varphi_{ix_1} L^K \varphi_{ix_1}
\]

\[
= \int \varphi_{ix_1} L^K \left( x_1 u_i - \sum_{l=1}^{m} a_{ilx_1} u_l \right)
\]

\[
= \int \varphi_{ix_1} \left( \lambda_i x_1 u_i - 2 \sum_{q=1}^{k} L^{k-q} X_1 L^{q-1} u_i \right)
\]

\[
= \lambda_i \int \varphi_{ix_1}^2 - 2 \int \left( \sum_{q=1}^{k} L^{k-q} X_1 L^{q-1} - 1 u_i \right) \varphi_{ix_1}, \quad i = 1, \ldots, m.
\]

Introducing a parameter $\beta$, we have

\[
(\lambda_{m+1} - \beta) \sum_i \int \varphi_{ix_1}^2 \leq \sum_i (\alpha_i - \beta) \int \varphi_{ix_1}^2 - 2 \sum_{i,q} \int L^{k-q} X_1 L^{q-1} u_i \cdot \varphi_{ix_1}.
\]

Let

\[
S = \sum_{i=1}^{m} \sum_{j=1}^{n} \int \left( \varphi_{ijx_j}^2 + \varphi_{ijy_j}^2 \right),
\]

\[
I_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} \left( \beta - \lambda_j \right) \int \left( \varphi_{ijx_j}^2 + \varphi_{ijy_j}^2 \right),
\]

\[
I_2 = -2 \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{q=1}^{k} \int \left( \varphi_{ijx_j} L^{k-q} X_j L^{q-1} u_i + \varphi_{ijy_j} L^{k-q} X_j L^{q-1} u_i \right).
\]
we have

\[(\lambda_{m+1} - \beta)S + I_1 \leq I_2.\]

Applying Lemma 2.2 yields

\[
m \leq \left[ \sum_{i=1}^{m} \int \left( \varphi_{ix_i}^2 + \varphi_{iy_j}^2 \right) \right]^{\frac{1}{2}} \left[ \sum_{i=1}^{m} \int \left( (X_j u_i)^2 + (Y_j u_i)^2 \right) \right]^{\frac{1}{2}}
\]

and summing over \(j\) gives

\[2mn \leq \delta I_1 + \frac{1}{\delta} \sum_{i=1}^{m} \frac{1}{\beta - \lambda_i} \int |\nabla L u_i|^2,\]

where \(\delta\) is a positive number. Then we have by Corollary 2.1

\[2mn \leq \delta I_1 + \frac{1}{\delta} \left( \sum_{i=1}^{m} \frac{\lambda_i^{\frac{1}{p}}}{\beta - \lambda_i} \right).\]

After choosing the minimum of the right-hand side, we derive a lower bound of \(I_1\):

\[(5.4) \quad I_1 \geq m^2 n^2 \left( \sum_{i=1}^{m} \frac{\lambda_i^{\frac{1}{p}}}{\beta - \lambda_i} \right)^{-1}.\]

Now we estimate \(I_2\). Notice the following relation by (2.1)

\[- \sum_{i,l,q} a_{ilx_j} \int u_l L^{q-1} X_j L^{q-1} u_i = \frac{1}{2} \sum_{i,l} a_{ilx_j} \int \left( u_l L^k (x_j u_i) - x_j u_i L^k u_i \right) = 0\]

and

\[- \sum_{i,l,q} a_{ily_j} \int u_l L^{q-1} Y_j L^{q-1} u_i = 0, \quad i, l = 1, \ldots, m, q = 1, \ldots, k.\]

Therefore we have

\[(5.5) \quad I_2 = -2 \sum_{i,j,q} \int \left( (x_j u_i L^{k-q} X_j L^{q-1} u_i + y_j u_i L^{k-q} Y_j L^{q-1} u_i) \right)\]

\[= -2 \sum_{i,j,q} \int \left( L^{k-q} (x_j u_i) X_j L^{q-1} u_i + L^{k-q} (y_j u_i) Y_j L^{q-1} U_i \right)\]

\[= -2 \sum_{i,j,q} \left\{ \int \left( (x_j L^{k-q} u_i - 2 \sum_{r=1}^{k-q-1} L^{k-q-r} X_j L^{r-1} u_i) X_j L^{q-1} u_i \right) \right.\]

\[- 2X_j L^{k-q-1} u_i \cdot X_j L^{q-1} u_i \right\]
\[ + \int \left( y_j L^{k-q} u_i - 2 \sum_{r=1}^{k-q-1} L^{k-q-r} Y_j L^{q-1} u_i - 2 Y_j L^{k-q-1} u_i \cdot Y_j L^{q-1} u_i \right) \].

Since
\[ \sum_{q=1}^{k} \int \left( \frac{x_j}{y_j} \right) L^{k-q} u_i \cdot \left( \frac{X_j}{Y_j} \right) L^{q-1} u_i \]
\[ = - \sum_{q=1}^{k} \left[ \int L^{k-q} u_i \cdot L^{q-1} u_i + \left( \frac{x_j}{y_j} \right) L^{k-q} u_i \cdot L^{q-1} u_i \right] \]
we have
\[ \sum_{q=1}^{k} \int \left( \frac{x_j}{y_j} \right) L^{k-q} u_i \cdot \left( \frac{X_j}{Y_j} \right) L^{q-1} u_i \]
\[ = - \frac{1}{2} \sum_{q=1}^{k} \int L^{k-q} u_i \cdot L^{q-1} u_i \]
\[ = - \frac{k}{2} \int L^{k-1} u_i \cdot u_i. \]

On the other hand
\[ \sum_{j} \int \left( X_j L^{k-q-1} U_i \cdot X_j L^{q-1} U_i + Y_j L^{k-q-1} u_i \cdot Y_j L^{q-1} u_i \right) \]
\[ = \int L^{k-1} u_i \cdot u_i \]
so it follows
(5.6)
\[ I_2 = (2n + 4) k \sum_{i=1}^{m} \int L^{k-1} u_i \cdot u_i \]
\[ + 4 \sum_{i,j,q} \int \left( \sum_{r=1}^{k-q-1} L^{k-q-r} X_j L^{r-1} u_i \cdot X_j L^{q-1} u_i \right) \]
\[ + \sum_{r=1}^{k-q-1} L^{k-q-r} Y_j L^{r-1} u_i \cdot Y_j L^{q-1} u_i \]
\[ = (2n + 4) k \sum_{i=1}^{m} \int L^{k-1} u_i \cdot u_i \]
\[ + 4 \sum_{i,j,q} \sum_{r=1}^{k-q} \left[ \sum_{s=0}^{k-q-r} C_{k-r-q}^{s}(-1)^{\frac{r+1}{2}} A_j^s \right] \]
\[ \cdot L^{k-q-r-s}(2T)^s L^{r-1} u_i \cdot X_j L^{q-1} u_i \]
\[
\begin{align*}
&\quad + \sum_{s=0}^{k-q-r} C_{k-q-r}^s (-1)^{j_s} A_j^{s+1} L^{k-q-r-s} (2T)^s L^{r-1} u_i \cdot Y_j L^{q-1} u_j \\
&= (2n+4) k \sum_{i=1}^{m} L^{k-1} u_i \cdot u_i + 4 \sum_{i,j,q} \sum_{r=1}^{k-q-1} (I_3 + I_4)
\end{align*}
\]

where we have used Lemma 2.4. We obtain that

\begin{equation}
I_{3(s \text{ odd})} + I_{4(s \text{ odd})}
\end{equation}

\begin{align*}
&= \sum_{s \leq k-q-r, s \text{ odd}} \left[ \int C_{k-q-r}^s (-1)^{j_s} Y_j L^{k-q-r-s} (2T)^s L^{r-1} u_i \cdot Y_j L^{q-1} u_i \\
&\quad + \int C_{k-q-r}^s (-1)^{j_s} X_j L^{k-q-r-s} (2T)^s L^{r-1} u_i \cdot Y_j L^{q-1} u_i \right] \\
&= - \sum_{s \leq k-q-r, s \text{ odd}} \left[ \int C_{k-q-r}^s (-1)^{j_s} (Y_j X_j - X_j Y_j) \right. \\
&\quad \cdot L^{k-q-r-s} (2T)^s L^{r-1} u_i \cdot L^{q-1} u_i \\
&= - \sum_{s \leq k-q-r, s \text{ odd}} 2^s C_{k-q-r}^s (-1)^{j_s} \int T^{s+1} L^{k-q-s-1} u_i \cdot L^{q-1} u_i \\
&= - \sum_{s \leq k-q-r, s \text{ odd}} 2^s C_{k-q-r}^s (-1)^{j_s} \int T^{s+1} u_i \cdot L^{k-s-2} u_i
\end{align*}

where \( T = Y_j X_j - X_j Y_j \). Similarly,

\begin{equation}
I_{3(s \text{ even})} + I_{4(s \text{ even})} = \sum_{s \leq k-q-r, s \text{ even}} 2^s C_{k-q-r}^s (-1)^{j_s} \int T^s u_i \cdot L^{k-s-1} u_i.
\end{equation}

First let \( k \) be odd. If \( s \) is odd, then by (2.14)

\begin{equation}
(5.9) \quad \int T^{s+1} u_i \cdot L^{k-s-2} u_i \\
= (-1)^{s+1/2} \int \left( T \right)^{s+1} L^{k-s-2} u_i)^2 \\
\leq \frac{1}{2(2n-1)^{s+1/2}} \left[ \int L^{k-s-1} u_i \cdot L^{k-s-1} u_i + \int L^{k-s-1} u_i \cdot L^{k-s-1} u_i \right] \\
= \frac{1}{2(2n-1)^{s+1/2}} \left[ \int L^k u_i \cdot u_i + \int L^{k-2} u_i \cdot u_i \right],
\end{equation}
and if \( s \) is even, then

\[
(5.10) \quad (-1)^{\left\lfloor \frac{s}{2} \right\rfloor} \int T^s u_i \cdot L^{k-s} u_i
\]

\[
= (-1)^{\left\lfloor \frac{s}{2} \right\rfloor + \left\lfloor \frac{s}{2} \right\rfloor} \left( T^s \cdot L^{s-1} \right)^2 u_i
\]

\[
\leq \frac{1}{2(2n-1)^{\frac{s}{2}}} \left[ \int L^{k-1} u_i \cdot L^{k-1} u_i + \int L^{k-2} u_i \cdot L^{k-2} u_i \right]
\]

\[
= \frac{1}{2(2n-1)^{\frac{s}{2}}} \left[ \int L^k u_i \cdot u_i + \int L^{k-2} u_i \cdot u_i \right] (k \geq 3).
\]

Summing (5.9) and (5.10) yields

\[
(5.11)
\]

\[
I_2 \leq (2n + 4) k \sum_{i=1}^{m} \int L^{k-1} u_i \cdot u_i
\]

\[
+ \frac{4}{2(n-1)^{\frac{s}{2}}} \sum_{i,j,q} \sum_{r=1}^{k-q-1} \left[ \sum_{s \leq k-q-r(s \text{ odd})} 2^s C_{k-q-r} 1 \right] \left( \int L^k u_i \cdot u_i + \int L^{k-2} u_i \cdot u_i \right)
\]

\[
+ \sum_{s \leq k-q-r(s \text{ even})} 2^s C_{k-q-r} \frac{1}{2(2n-1)^{\frac{s}{2}}} \left( \int L^k u_i \cdot u_i + \int L^{k-2} u_i \cdot u_i \right)
\]

\[
\leq (2n + 4) k \sum_{i=1}^{m} \int L^{k-1} u_i \cdot u_i
\]

\[
= 2 \sum_{i,j,q} \sum_{r=1}^{k-q-1} \left( \sum_{s \leq k-q-r(s \text{ odd})} 2^s C_{k-q-r} \frac{1}{2(n-1)^{\frac{s}{2}}} \right) + \sum_{s \leq k-q-r(s \text{ even})} 2^s C_{k-q-r} \frac{1}{(2n-1)^{\frac{s}{2}}}
\]

\[
\cdot \left( \int L^k u_i \cdot u_i + \int L^{k-2} u_i \cdot u_i \right)
\]

\[
\leq (2n + 4) k \sum_{i=1}^{m} \left( \int L^k u_i \cdot u_i \right) \frac{k-1}{k}
\]

\[
+ C_1(n,k) \sum_{i=1}^{m} \left[ \lambda_i + \left( \int L^k u_i \cdot u_i \right) \frac{k-1}{k} \right]
\]

\[
\leq (2n + 4) k \sum_{i=1}^{m} \lambda_i \frac{k-1}{k} + C_1(n,k) \sum_{i=1}^{m} \left[ \lambda_i + \lambda_i \frac{k-1}{k} \right],
\]
where we have used Lemma 2.3. Note that
\[ C_1(n, k) \leq 2n \sum_{q=1}^{k} \sum_{r=1}^{k-q-1} \left( 1 + \frac{2}{(2n-1)^q} \right)^{k-q-r}. \]

Substituting (5.11) in (5.3) leads to
\[
(\lambda_{m+1} - \beta)s \leq (2n + 4)k \sum_{i=1}^{m} \frac{\lambda_{s}^{k-1}}{k} \\
+ C_1(n, k) \sum_{i=1}^{m} \left( \lambda_i + \frac{\lambda_{s}^{k-1}}{k} \right) - m^2 n^2 \left( \sum_{i=1}^{m} \frac{1}{\beta - \lambda_i} \right)^{-1}.
\]

Along with the method in [10], we obtained (5.1).

Now let \( k \) be even. If \( s \) is odd, then by Corollary 2.2
\[
(-1)^{\lfloor \frac{s}{2} \rfloor + 1} \int T^{s+1} u_i L^{k-s-2} u_i = \int T^{s+1} L^{k-s-3} u_i \cdot T^{s+1} L^{k-s-3} + 1 u_i \\
\leq \frac{1}{(2n-1)^{s+1}} \int L^{k-s-1} u_i L^{k-s-2} u_i \\
= \frac{1}{(2n-1)^{s+1}} \int L^{k-1} u_i \cdot u_i
\]
and if \( s \) is even, then
\[
(-1)^{\lfloor \frac{s}{2} \rfloor} \int T^s u_i L^{k-s-1} u_i = \int T^s L^{k-s-2} u_i \cdot T^s L^{k-s-2} + 1 u_i \\
\leq \frac{1}{(2n-1)^{s}} \int L^{k-s-2} u_i L^{k-s-2} + 1 u_i \\
= \frac{1}{(2n-1)^{s}} \int L^{k-1} u_i \cdot u_i.
\]
Summing two inequalities above yields
\[
I_2' \leq (2n + 4)K \sum_{i=1}^{m} \int L^{k-1} u_i \cdot u_i \\
+ 4 \sum_{i,j,q} \sum_{r=1}^{k-q-1} \left( \sum_{s \leq k-q-r(s \text{ odd})} \frac{2^s C_{k-q-r}^s}{(2n-1)^{s+1}} \\
+ \sum_{s \leq k-q-r(s \text{ even})} \frac{2^s C_{k-q-r}^s}{(2n-1)^{s}} \right) \int L^{k-1} u_i \cdot u_i \\
\triangleq (2n + 4)K \sum_{i=1}^{m} \int L^{k-1} u_i \cdot u_i + C_2(n, k) \sum_{i=1}^{m} \int L^{k-1} u_i \cdot u_i \\
\leq (2n + 4)K \sum_{i=1}^{m} \lambda_i^{k-1} + C_2(n, k) \sum_{i=1}^{m} \lambda_i^{k-1}.
\]

Note that
\[
C_2(n, k) \leq 4n \sum_{q=1}^{k} \sum_{r=1}^{k-q-1} \left( 1 + \frac{2}{(2n-1)^{q}} \right)^{k-q-r}.
\]

Then proceeding with the same way for proving (5.1) we deduce (5.2). \(\square\)


Consider the eigenvalue problem
\[
\begin{aligned}
-L u = \lambda u, & \quad \text{in } \Omega \\
 u = 0 & \quad \text{on } \Omega
\end{aligned}
\]
where \(L\) denote the generalized Baouendi-Grushin operator
\[
L = \sum_{i=1}^{M} X_i^2 + \sum_{j=1}^{N} Y_j^2
\]
with the nonsmooth vector fields
\[
X_j = \frac{\partial}{\partial x_i}, \quad i = 1, \ldots, M, \quad Y_j = |x|^\alpha \frac{\partial}{\partial y_j}, \quad j = 1, \ldots, N,
\]
\(\alpha \geq 1, \alpha \in R, M, N \geq 1\) and \(M + N = n\). \(\Omega\) is a bounded domain in \(R^n\) with smooth boundary \(\partial \Omega\).

The existence of eigenvalues for the problem (6.1) can be proved with the method in [12] and the embedding theorem in [6] and then (6.1) possesses a system of eigenfunctions \(\{u_h\}\) that forms an orthonormal base, with the corresponding eigenvalues \(\{\lambda_h\} : 0 < \lambda_1 \leq \lambda_2 \leq \ldots\).
By choosing the trial functions

\[ \varphi_{hx_i} = x_i u_h - \sum_{i=1}^{m} a_{hix_j} u_t, \quad \varphi_{hy_i} = y_i u_h - \sum_{i=1}^{m} a_{hy_j} u_t, \quad h = 1, \ldots, m, \]

we have

\[ (\lambda_{m+1} - \lambda_m) \sum_{h=1}^{m} \left( \sum_{i=1}^{M} \int_{\Omega} \varphi_{hx_i}^2 + \sum_{j=1}^{N} \int_{\Omega} \varphi_{hy_j}^2 \right) \]

\[ \leq -2 \sum_{i=1}^{m} \left( \sum_{i=1}^{M} \varphi_{hx_i} X_i u_h + \sum_{j=1}^{N} \int |x|^\alpha \varphi_{hy_j} Y_j u_h \right). \]

Noting that

\[ \sum_{i=1}^{M} \sum_{h=1}^{m} \int \varphi_{hx_i} X_i u_h = -\frac{mM}{2}, \]

\[ \sum_{j=1}^{N} \sum_{h=1}^{m} \int \varphi_{hy_j} Y_j u_h = -\frac{N}{2} \sum_{h=1}^{m} \int |x|^{2\alpha} u_h^2, \]

and Schwarz’s inequality, we get

\[ \frac{mM}{2} \leq \frac{mM}{2} + \frac{N}{2} \sum_{h=1}^{m} \int |x|^{2\alpha} u_h^2 \]

\[ \leq \sum_{h=1}^{m} \left[ \sum_{i=1}^{M} \int |\varphi_{hx_i} X_i u_h| + \sum_{j=1}^{N} \int |x|^\alpha |\varphi_{hy_j} Y_j u_h| \right] \]

\[ \leq \max(1, d^\alpha) \left[ \sum_{h=1}^{m} \left( \sum_{i=1}^{M} \int |\varphi_{hx_i}|^2 \right) + \sum_{j=1}^{N} \int |\varphi_{hy_j}|^2 \right]^{\frac{1}{2}} \]

\[ \cdot \left[ \sum_{h=1}^{m} \int (-L u_h) u_h \right]^{\frac{1}{2}} \]

where \( d \) is the diameter of \( \Omega_y \), the projection of \( \Omega \) in the \( y \)-space. Then

\[ \sum_{h=1}^{m} \left( \sum_{i=1}^{M} \int |\varphi_{hx_i}|^2 + \sum_{j=1}^{N} \int |\varphi_{hy_j}|^2 \right) \geq \frac{m^2 M^2}{4 \max(1, d^{2\alpha}) \sum_{h=1}^{m} \lambda_h}. \]

Returning to (6.5), we prove the following:
Theorem 6.1. Let $m \geq 1$, then
\[
\lambda_{m+1} = \lambda_m \leq \frac{4n}{mM^2} \max(1, d^{4\alpha}) \sum_{h=1}^{m} \lambda_h.
\]

Remark 6.1. This shows that the upper bounds for eigenvalues of (6.1) depends on the domain.

Acknowledgments. This research was supported by the National Natural Science Foundation of China. The authors would like to thank Professor Luo Xuebo for his hospitality. They also would like to thank the referees for helpful comments and suggestions.

References


Received October 22, 1999 and revised March 29, 2002. This research was supported by the National Natural Science Foundation of China.

DEPARTMENT OF APPLIED MATHEMATICS
NORTHWESTERN POLYTECHNICAL UNIVERSITY
XI‘AN, SHAANXI, 710072
PEOPLE’S REPUBLIC OF CHINA
E-mail address: taoer558@sohu.com

DEPARTMENT OF APPLIED MATHEMATICS
NORTHWESTERN POLYTECHNICAL UNIVERSITY
XI‘AN, SHAANXI, 710072
PEOPLE’S REPUBLIC OF CHINA
PROOF OF THE DOUBLE BUBBLE CONJECTURE IN $\mathbb{R}^4$
AND CERTAIN HIGHER DIMENSIONAL CASES

Ben W. Reichardt, Cory Heilmann, Yuan Y. Lai, and Anita Spielman

We prove that the standard double bubble is the minimizing double bubble in $\mathbb{R}^4$ and in certain higher dimensional cases, extending the recent work in $\mathbb{R}^3$ of Hutchings, Morgan, Ritoré and Ros.

1. Introduction.

1.1. The Double Bubble Conjecture.

Conjecture 1.1 (Double Bubble Conjecture). The least-area hypersurface enclosing and separating two given volumes in $\mathbb{R}^n$ is the standard double soap bubble of Figure 1, consisting of three $(n-1)$-dimensional spherical caps intersecting at 120 degree angles. (For the case of equal volumes, the middle sphere is a flat disk.)

In 1990, Foisy et al. [F] proved the Double Bubble Conjecture in $\mathbb{R}^2$. In 1995, Hass, Hutchings and Schlafly [HHS], [HS] used a computer to prove the conjecture for the case of equal volumes in $\mathbb{R}^3$. Most recently, in 2000, Hutchings, Morgan, Ritoré and Ros [HMRR] have used stability arguments to prove the conjecture for all cases in $\mathbb{R}^3$. Morgan’s reference [M] discusses these results.

Here, we extend the methods of Hutchings et al. to higher dimensions. Component bounds after Hutchings [H] guarantee that the “1 + $k$” double

Figure 1. The standard double bubble, consisting of three spherical caps meeting at 120 degree angles, is the conjectured least-area hypersurface that encloses two given volumes in $\mathbb{R}^n$. 
bubble — a double bubble in which one region is connected and the other region has $k$ components — is the only alternative to the standard double bubble as the minimizing hypersurface in $\mathbb{R}^4$ and for sufficiently unequal volumes in $\mathbb{R}^n$, and that the larger region is connected (2.2, 2.5). By showing such bubbles unstable, we prove the Double Bubble Conjecture in $\mathbb{R}^4$ (Theorem 9.1) and, when the larger region has more than $2/3$ the total volume, in $\mathbb{R}^n$ (Theorem 9.2).

1.2. The Instability Argument. An area-minimizing double bubble $\Sigma$ exists and has an axis of rotational symmetry $L$. Consider small rotations about a line $M$ orthogonal to $L$, chosen such that the points of tangency between the rotation vector field $v$ and $\Sigma$ separate the bubble into at least four pieces. Then we can linearly combine the restrictions of $v$ to each piece to obtain a vector field which vanishes on one piece and preserves volume. By regularity for eigenfunctions, $v$ is tangent to certain related parts of $\Sigma$, implying that they are spheres centered on $L \cap M$. This is the instability argument of [HMRR] behind Theorem 4.1. The corollaries to the theorem use the spherical pieces of $\Sigma$ to show that it must be the standard double bubble.

We consider $\Sigma$ a nonstandard double bubble with the larger region connected, and assume that $\Sigma$ is a minimizer. In §6 and §7, we look at isolated parts of $\Sigma$ — its “root” and its “leaves” according to an associated tree structure — and we classify all possible root and leaf configurations in which no useful perturbation axis $M$ can be found. In §8 we combine our local classification results to nevertheless find a suitable $M$. By the above argument, $\Sigma$ cannot in fact be a minimizer. So, for example, the double bubble with cross-section as in Figure 2 cannot be a minimizer.

Having eliminated all nonstandard double bubbles from consideration, the only possible minimizer left is the standard double bubble.

1.3. Open Questions. It can be shown that the leaf classification of Proposition 7.1 remains valid without the restriction that the larger region be connected. The instability of all nonstandard double bubbles in which one region is connected follows.

However, our component bounds are not strong enough to assure that one region must always be connected; for higher dimensions, they in general only establish that the larger region has at most 3 components and the smaller region has a finite number of components [HLRS]. But we have not been able to prove the instability of all double bubbles where both regions are disconnected.

Indeed, our methods fail to prove unstable the $2 + 2$ double bubble generated by rotating the curves of Figure 3 about the symmetry axis, in $\mathbb{R}^5$ or higher dimensions (§5 explains the rotation numbers attached to the vertices in the figure). Showing that this configuration is not minimizing, together with our bounds in [HLRS] and the results here, would prove the Double Bubble Conjecture in $\mathbb{R}^5$. 
Figure 2. The lines orthogonal to Σ through the points of the separating set all pass through $M$. Σ cannot be a minimizer.

Figure 3. A $2 + 2$ double bubble might not have any disallowed interior separating sets, thus cannot be eliminated as unstable by our methods.

2. Double bubbles and component bounds.

A double bubble is a piecewise smooth oriented hypersurface $\Sigma \subset \mathbb{R}^n$ consisting of three compact pieces $\Sigma_1$, $\Sigma_2$ and $\Sigma_0$ (smooth on their interiors), with a common boundary such that $\Sigma_1 \cup \Sigma_0$, $\Sigma_2 \cup \Sigma_0$ enclose two regions of given volumes. Let $A_n(v, w)$ be the least area of a double bubble enclosing regions $R$ of volume $v$ and $S$ of volume $w$. Let $\tilde{A}_n(v, w) \geq A_n(v, w)$ be the area of the standard double bubble. Let $A_n(v) = n\pi^{1/2}v^{n-1}/(n/2)!^{1/n}$ be the area of a sphere of volume $v$.

[H] shows strict concavity of the least area function $A_n(v, w)$ and uses it to find bounds on the number of components of minimizing double bubbles. We will list some of his results in $\mathbb{R}^n$ and numerically compute them in $\mathbb{R}^4$. (See [HLRS] for more extensive numerical computations.)
Theorem 2.1 ([H, Theorem 3.2]). If \( n \geq 3 \), if \((v_1, w_1), (v_2, w_2)\) are two pairs of nonnegative volumes, and if \(0 < t < 1\), then
\[
A_n(tv_1 + (1-t)v_2, tw_1 + (1-t)w_2) > tA_n(v_1, w_1) + (1-t)A_n(v_2, w_2).
\]

Theorem 2.1 yields a dimension-independent component bound for unequal volumes:

Corollary 2.2 ([H, Theorem 3.5]). If \( v > 2w \), then in any least-area enclosure of volumes \( v, w \) in \( \mathbb{R}^n \), \( \mathbb{R}^n \) the region of volume \( v \) is connected.

A slightly more sophisticated decomposition argument, together with the pigeonhole principle, gives a better component bound:

Theorem 2.3 ([H, Theorem 4.2]). Consider a minimal enclosure of volumes \( v, w \) in \( \mathbb{R}^n \). Then the number of components \( k \) of \( \mathbb{R}^n \) the region of volume \( v \) satisfies
\[
2A_n(v, w) \geq A_n(w) + A_n(v + w) + A_n(v) \cdot k^{1/n}.
\]

Clearly, \( k \) in Theorem 2.3 is finite:

Corollary 2.4 ([H, Corollary 4.3]). A minimal enclosure of two volumes in \( \mathbb{R}^n \) separates \( \mathbb{R}^n \) into finitely many components.

Proposition 2.5. In a minimizing double bubble in \( \mathbb{R}^4 \), a region of at least half the total volume is connected.

Proof. By concavity Theorem 2.1, for \( v \in [0, 1] \),
\[
A_4(v, 1-v) \leq A_4(0.5, 0.5) \leq \tilde{A}_4(0.5, 0.5) = \left( \frac{4}{3} + \frac{3\sqrt{3}}{4\pi} \right)^{1/4},
\]
by computation. Hence, letting \( w = 1 - v \) in Theorem 2.3, we obtain
\[
k^{1/4} \leq \frac{1}{v^{3/4}} \left( \left( \frac{64}{3} + \frac{12\sqrt{3}}{\pi} \right)^{1/4} - 1 \right) - \left( \frac{1}{v} - 1 \right)^{3/4} < \frac{1.3}{v^{3/4}} - \left( \frac{1}{v} - 1 \right)^{3/4} \quad \Rightarrow \quad b(v).
\]
Differentiating \( b(v) \) gives that \( b'(v) \) has the same sign as \( 1 - 1.3(1 - v)^{1/4} \), which is increasing. Hence \( b'(v) \) passes from having negative sign to having positive sign, so on a closed interval \( b(v) \) attains its maximum at an endpoint. Now,
\[
b(0.5)^4 = (1.3 \cdot 2^{3/4} - 1)^4 < 1.99
\]
\[
b(2/3)^4 = (1.3 \cdot (3/2)^{3/4} - (1/2)^{3/4})^4 < 1.86.
\]
Hence \( b(v)^4 < 2 \) for \( v \in [0.5, 2/3] \), implying that \( k = 1 \); \( R \) is connected. By Corollary 2.2, \( R \) is connected for \( v > 2/3 \). Hence \( R \) is connected for all \( v \in [0.5, 1] \).
3. Structure of minimal double bubbles.

The work of Almgren ([A] and see [M, Chapt. 13]) tells us that an area-minimizing double bubble enclosing any two given volumes in $\mathbb{R}^n$ exists and is almost everywhere regular, if we allow disconnected regions. It is both stationary and stable. Hence, each region has a well-defined pressure, positive by [H, Corollary 3.3].

Lemma 3.1 ([HMRR, Lemma 6.4]). In a minimizing double bubble for unequal volumes, the smaller region has larger pressure.

This result follows easily from Hutchings concavity Theorem 2.1. Hutchings further classifies possible nonstandard minimizing double bubbles:

Theorem 3.2 ([H, Theorem 5.1]). Any nonstandard minimal double bubble is a hypersurface of revolution about some line $L$, composed of pieces of constant mean curvature hypersurfaces meeting in threes at 120 degree angles. The bubble is a topological sphere with a tree $T$ of annular bands attached, as in Figure 4. The two caps of the bottom component are pieces of spheres, and the root of the tree has just one branch.

Hence, any minimal double bubble is determined by an upper half-planar diagram of arcs of generating curves which, when rotated about $L$, generate the double bubble. By studying these generating curves, we will eliminate as unstable nonstandard double bubbles.

[HY] shows that the only constant mean-curvature hypersurfaces of revolution are Delaunay hypersurfaces (Figure 5 and see [D], [E]):
Figure 5. Smooth regions of the cluster are parts of Delaunay hypersurfaces: Catenoid, nodoid, unduloid, vertical plane, sphere.

**Theorem 3.3** ([HMRR, Proposition 4.3]). Let $\Gamma$ be a complete upper half-planar generating curve which, when rotated about $L$, generates a hypersurface $\Sigma$ with constant mean curvature. Then exactly one of the following statements holds:

1) $\Gamma$ is a curve of catenary type and $\Sigma$ is a hypersurface of catenoid type.
2) $\Gamma$ is a locally convex curve and $\Sigma$ is a nodoid.
3) $\Gamma$ is a periodic graph over $L$ and $\Sigma$ is an unduloid or a cylinder.
4) $\Gamma$ is a ray orthogonal to $L$ and $\Sigma$ is a vertical hyperplane.
5) $\Gamma$ is a semi-circle and $\Sigma$ is a sphere.

The Delaunay hypersurfaces with nonzero mean curvature are the sphere, unduloid and nodoid. If $\Sigma$ has positive mean curvature upward then it must be a nodoid. If $\Gamma$ is not graph, then $\Sigma$ must be either a nodoid or a hyperplane.

**4. Instability by separation.**

Let $\Sigma \subset \mathbb{R}^n$ be a regular stationary double bubble of revolution about axis $L$, with upper half planar generating curves $\Gamma$ consisting of arcs $\Gamma_i$, with interiors $\Gamma_i$, ending either at the axis or in threes at vertices $v_{ijk}$.

We consider the map $f : \bigcup \Gamma_i \longrightarrow L \cup \{\infty\} \equiv [-\infty, +\infty]/(-\infty \sim +\infty)$ which maps each $p \in \bigcup \Gamma_i$ to the point $L(p) \cap L$, where $L(p)$ denotes the normal line to $\Gamma$ at $p$. Later we will denote the limiting values of $f$ on the left and right endpoints of $\Gamma_i$ by $iA, iB \in [-\infty, +\infty]$, respectively; for consistency, we will often simply consider $f(p)$ as its preimage in $[-\infty, +\infty]$.

(With this notation, if $iA \in f(\Gamma_j)$ and $\Gamma_j$ is not a circle or hyperplane, then for all $p \in \Gamma_i$ sufficiently close to the left endpoint, $f(p) \in f(\Gamma_j)$.)

**Theorem 4.1** ([HMRR, Proposition 5.2]). Consider a stable double bubble of revolution $\Sigma \subset \mathbb{R}^n$, $n \geq 3$, with axis $L$. Assume that there is a minimal set of points $\{p_1, \ldots, p_k\}$ in $\bigcup \Gamma_i$ with $f(p_1) = \cdots = f(p_k)$ which separates $\Gamma$. 
Then every connected component of $\Sigma$ which contains one of the points $p_i$ is part of a sphere centered at $x$ (if $x \in L$) or part of a hyperplane orthogonal to $L$ (in the case $x = \infty$).

We sketched the proof of this theorem in our introduction §1.2.

**Corollary 4.2.** No generating arc which turns downward past the vertical can have an internal separating set, i.e., two points $p_1 \neq p_2$ in the arc, with $f(p_1) = f(p_2)$.

*Proof.* Otherwise, by Theorem 4.1, the arc would have to be part of either a circle with center on the axis $L$ or a line perpendicular to $L$. But neither turns past the vertical. □

**Corollary 4.3.** No generating arc which is not part of a vertical line can go vertical twice, including at least once in its interior.

*Proof.* Such an arc (a nodoid by Theorem 3.3) has a separating set $\{f^{-1}(x)\}$ for some $x$ with $|x|$ large enough, contradicting Corollary 4.2. □

**Corollary 4.4.** Consider a nonstandard minimizing double bubble. Then there is no $x \in L \cup \{\infty\}$ such that $f^{-1}(x) - \{\text{two circular caps}\}$ contains points in the interiors of distinct $\Gamma_i$ which separate $\Gamma$.

*Proof.* For $x \in L$, the statement is [HMRR, Proposition 5.7]. Arguments using “force balancing” show that more pieces of the minimizer are spherical and hence the bubble is the standard double bubble. For $x = \infty$, note that a separating set crosses at least one outer boundary. By Theorem 4.1, this boundary is a vertical line, contradicting positive pressure of the regions. □

We will consider various nonstandard double bubbles, and show that they violate one of the above corollaries of Theorem 4.1, hence cannot be minimizing.

5. **Rotation notation.**

A nonstandard minimal double bubble’s generating curves can be further classified by how many notches $m$ a vertex has been rotated from the standard position of Figure 6 in which all the generating curves are graphs (unlike Figure 4). A positive rotation notch about a vertex corresponds to an arc passing the vertical counterclockwise, as occurs for $\Gamma_3$ from Figure 6 to Figure 7, and from Figure 8(a) to 8(b). The extreme position with an arc leaving the vertex at the vertical divides two consecutive $m$ cases. If the limiting value of $f$ along the vertical arc is $+\infty$ (or the arc is a straight line), the position is assigned the smaller $m$ rotation number, and if the limiting value is $-\infty$, the position is given the larger $m$ value.

The rotation numbers for our earlier 4 + 4 example are indicated in Figure 9.
Figure 6. If all the generating curves are graphs, then $m = 0$ for each vertex.

Figure 7. From the curves of Figure 6, vertex $v_{123} \equiv \Gamma_1 \cap \Gamma_2 \cap \Gamma_3$ has turned one counterclockwise “notch,” since $\Gamma_3$ has passed the vertical. Hence $m = 1$ for $v_{123}$.

Figure 8. A close-up of $v_{123}$ as it turns one notch counterclockwise. In (a), $m = 0$ and $2A < 1B < 3A$; on the right, $3A = +\infty$. In (b), $m = 1$ and $3A < 2A < 1B$; $3A = -\infty$ on the right.

The “root” of a nonstandard minimizing double bubble corresponds to the root of its associated tree of Theorem 3.2 and Figure 4. The root involves five arcs including circular caps to either side, as in Figure 10.

**Proposition 6.1.** In a minimizer, consider a root with the notation of Figure 10. Then \((m_1, m_2) \in \{(-1, -1), (0, -1), (0, 0), (1, -1), (1, 0), (1, 1)\}\).

**Proof.** \(\Gamma_1\) and \(\Gamma_5\) are parts of semi-circles, turning inward by positive pressure of the regions. It follows that \(m_1, m_2 \in \{-1, 0, 1\}\).

If \((m_1, m_2) = (-1, 0)\) as in Figure 11, then \([-\infty, 3B) \subset f(\Gamma_3)\), where \(3B\) denotes the image under \(f\) of the right-hand endpoint of \(\Gamma_3\). Consideration of vertex \(v_{345} = \bar{\Gamma}_3 \cap \bar{\Gamma}_4 \cap \bar{\Gamma}_5\) gives \(4B < 3B\). Hence \(4B \in f(\Gamma_3)\), giving a separating set through \(\Gamma_3\) and \(\Gamma_4\). Since this contradicts Corollary 4.4, \((m_1, m_2) \neq (-1, 0)\). Similarly, \((m_1, m_2) \neq (0, 1)\).

If \((m_1, m_2) = (-1, 1)\), then \(\Gamma_3\) goes vertical twice in its interior, violating Corollary 4.3. Hence \((m_1, m_2) \neq (-1, 1)\), as asserted. \(\square\)
Figure 11. If \((m_1, m_2) = (-1, 0)\), then \([-\infty, 3B) \subset f(\Gamma_3)\) and \(4B < 3B\), so \(4B \in f(\Gamma_3)\).

Figure 12. If \(\Gamma_2\) turns vertical downward after leaving the left circular cap, then the rotation number \(m_1\) of vertex \(v_{123}\) is either \(-1, 0\) or \(1\).

Proposition 6.2. In a minimizer, consider a root with the notation of Figure 10. Then neither \(\Gamma_2\) nor \(\Gamma_4\) can turn vertical downward after leaving a circular cap.

Proof. Suppose \(\Gamma_2\) turns vertical downward after leaving vertex \(v_{123}\). By Theorem 3.3 and positive pressure of the regions, \(\Gamma_2\) is a concave rightward nodoid. By Proposition 6.1, \(m_1 \in \{-1, 0, 1\}\), as shown in Figure 12.

First, consider \(m_1 = -1\). Then \((-\infty, 2A) \subset f(\Gamma_2)\). By Proposition 6.1, \(m_2 \in \{-1, 0, 1\}\), so \(\Gamma_3\) goes vertical before reaching \(v_{345}\). Hence \([-\infty, 3B) \subset f(\Gamma_3)\).

Second, consider \(m_1 = 0\). Then \((-\infty, 2A) \subset f(\Gamma_2)\) again, and \(3A < 2A\).

Third, consider \(m_1 = 1\). Then \(f(\Gamma_2) = L \cup \{\infty\}\).

In each case, \(f(\Gamma_2) \cap f(\Gamma_3) \neq \emptyset\), giving a separating set contrary to Corollary 4.4. Therefore \(\Gamma_2\) cannot turn vertical downward after leaving \(v_{123}\). Symmetrical considerations show that \(\Gamma_4\) cannot turn vertical downward after leaving \(v_{345}\). \(\square\)

7. Leaf stability.

A “leaf” of a nonstandard minimizing double bubble corresponds to a leaf of its associated tree of Theorem 3.2 and Figure 4. A leaf involves four arcs, with a standard notation as in Figure 13. We say that one case “models” another if they are symmetrical under horizontal reflection and/or relabelling.
Figure 13. A leaf involves four arcs: \( \Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4 \). In general each vertex can be rotated \( m_i \) notches counterclockwise from the pictured configuration, in which all arcs are graphs and \( m_1 = m_2 = 0 \).

Figure 14. The two near graph cases \((0,0)\) and \((0,1)\), shown here, and the cases \((0,2)\) and \((2,1)\) model all leaves belonging to the smaller region. Case \((0,0)\) models case \((3,3)\), and case \((0,1)\) models cases \((-2,-3)\), \((-1,0)\) and \((3,2)\), modulo \((6,6)\).

Figure 15. Case \((0,2)\) models cases \((-2,0)\), \((-1,-3)\) and \((3,1)\).

Case \((m_1,m_2) = (a,b)\) models cases \((a,b)\), \((-b,-a)\), \((b + 3, a + 3)\), and \((3 - a, 3 - b)\), up to rotation by \((6,6)\).

Proposition 7.1. In a minimizer for unequal volumes, consider a leaf belonging to the smaller region, with the notation of Figure 13. Then the only possible \((m_1,m_2)\) cases, up to rotation by \((6,6)\), are:

\[
(0, 0), (3, 3); (-2, -3), (-1, 0), (0, 1), (3, 2);
\]

\[
(-2, 0), (-1, -3), (0, 2), (3, 1); \quad \text{and} \quad (-2, -1), (-1, -2), (1, 2), (2, 1).
\]

In particular, every leaf is modeled on one of the four of Figures 14, 15 and 16.

Proof. First we need Lemma 7.2, in which we will use verticality arguments (refer to Corollary 4.3) to obtain general bounds on \((m_1,m_2)\) cases. We will then eliminate as unstable individual cases.
Figure 16. Case (2,1) models cases (−2, −1), (−1, −2) and (1, 2).

Lemma 7.2. With the assumptions as above, the only possibly stable rotation cases for \((m_1, m_2)\), up to rotation by \((6, 6)\), are:

- \(m_1 \in \{-2, -1\} \Rightarrow -3 \leq m_2 \leq 0\)
- \(m_1 = 0 \Rightarrow -3 \leq m_2 \leq 3\)
- \(m_1 \in \{1, 2\} \Rightarrow 0 \leq m_2 \leq 3\)
- \(m_1 = 3 \Rightarrow 0 \leq m_2 \leq 6\)

Proof. First, consider \(m_1 \in \{-2, -1\}\). \(\Gamma_3\) leaves \(v_{123}\) to the right, while \(\Gamma_2\) leaves \(v_{123}\) to the left. If \(m_2 = -4\), \(\Gamma_3\) enters \(v_{234}\) from the left, thus goes vertical twice (in its interior) if \(m_2 \leq -4\). If \(m_2 = 1\), \(\Gamma_2\) enters \(v_{234}\) from the right, thus goes vertical twice if \(m_2 \geq 1\). By Corollary 4.3, \(-3 \leq m_2 \leq 0\), as asserted.

Second, consider \(m_1 = 0\). \(\Gamma_3\) leaves \(v_{123}\) going upward to the right, while \(\Gamma_2\) leaves \(v_{123}\) going downward to the right. If \(m_2 = -4\), \(\Gamma_3\) enters \(v_{234}\) from the left, thus goes vertical twice if \(m_2 \leq -4\). If \(m_2 = 4\), \(\Gamma_2\) enters \(v_{234}\) from the left, thus goes vertical twice if \(m_2 \geq 4\). By Corollary 4.3, \(-3 \leq m_2 \leq 3\), as asserted.

Third, consider \(m_1 \in \{1, 2\}\). Considerations symmetrical to those of the cases \(m_1 \in \{-2, -1\}\) give \(0 \leq m_2 \leq 3\), as asserted.

Fourth, consider \(m_1 = 3\). Considerations symmetrical to those of the case \(m_1 = 0\) give \(0 \leq m_2 \leq 6\), as asserted. \(\square\)

To finish the proof of Proposition 7.1, we will use slightly more involved arguments to show that leaves with the following \((m_1, m_2)\) rotation pairs cannot belong to the smaller region of a minimizer:

\((-2, -2), (-1, -1), (0, -3), (0, -2), (0, -1), (0, 3), (1, 0), (1, 1), (1, 3), (2, 0), (2, 2), (2, 3), (3, 0), (3, 4), (3, 5), (3, 6)\).

Note that by Lemma 3.1, the leaf component has positive pressure larger than that of the adjacent component; \(\Gamma_2\) and \(\Gamma_3\) rotate about \(L\) to form hypersurfaces of positive mean curvature into the leaf.

- Case (1, 1) models cases \((-2, -2), (-1, -1), (2, 2)\); see Figure 17. \(4A \in f(\Gamma_3)\), since \(3B < 4A\) and \((3B, +\infty) \subset f(\Gamma_3)\).
Figure 17. Case (1, 1) models cases (−2, −2), (−1, −1), (2, 2).

Figure 18. Case (1, 0) models cases (0, −1), (2, 3), (3, 4). Case (2, 0) models cases (0, −2), (1, 3), (3, 5).

Figure 19. Case (3, 0) models cases (0, −3), (0, 3), (3, 6).

$\Gamma_2$ goes vertical, so by Theorem 3.3 must be part of a concave leftward nodoid. Hence the normal to $\Gamma_4$ at $v_{234}$ stays right of $\Gamma_2$ and in particular of $v_{123}$, implying $2A < 4A$. Since $(2A, +\infty] \subset f(\Gamma_2)$, $4A \in f(\Gamma_2)$.

Corollary 4.4 for $\Gamma_2$, $\Gamma_3$, $\Gamma_4$ implies instability.

- Case (1, 0) models cases (0, −1), (2, 3), (3, 4), and case (2, 0) models cases (0, −2), (1, 3), (3, 5); see Figure 18.

For both cases, again $3B < 4A$ and $(3B, +\infty] \subset f(\Gamma_3)$ imply $4A \in f(\Gamma_3)$. Since $4A < 2B$, Corollary 4.4 for $\Gamma_2$, $\Gamma_3$, $\Gamma_4$ yields $4A \leq 2A$.

Therefore, the net angle $\theta_3$ through which $\Gamma_3$ turns satisfies $\theta_3 > 180$ degrees, since $v_{123}$ is clearly left of $v_{234}$. Also, $\Gamma_3$ leaves $v_{234}$ above the horizontal, and Corollary 4.2 for $\Gamma_3$ gives $3A \leq 3B$. Since $\Gamma_2$ rotates about $L$ to form a hypersurface of positive mean curvature upwards into the leaf, by Theorem 3.3 $\Gamma_2$ is a (strictly convex) nodoid. We can now apply Lemma 7.3 to obtain $2A < 4A$, a contradiction.

- Case (3, 0) models cases (0, −3), (0, 3), (3, 6); see Figure 19.

$\Gamma_2$ goes vertical, so by Theorem 3.3 must be part of a concave rightward nodoid or a hyperplane, contradicting Lemma 3.1. $\Box$
Lemma 7.3 ([HMRR, Corollary 5.10]). Consider the (1,0), (2,0) and (2,1) cases of Figures 18 and 16. Assume that the net angle $\theta_3$ through which $\Gamma_3$ turns exceeds 180 degrees, that $\Gamma_3$ leaves $v_{234}$ at or above the horizontal, that $3A \leq 3B$, and that $\Gamma_2$ is strictly convex. Then $2A < 4A$.

Corollary 7.4. For a (2,1) leaf in a minimizer, as in Figure 16, $\Gamma_3$ leaves $v_{234}$ below the horizontal.

Proof. $4A \in f(\Gamma_3)$ since $3B < 4A$ and $(3B, +\infty) \subset f(\Gamma_3)$. Since also $(2A, +\infty) \subset f(\Gamma_2)$, Corollary 4.4 for $\Gamma_2, \Gamma_3, \Gamma_4$ yields $4A \leq 2A$.

$\Gamma_2$ goes vertical, so by Theorem 3.3 must be part of a concave leftward nodoid. Hence the normal to $\Gamma_4$ at $v_{234}$ stays right of $\Gamma_2$ and in particular of $v_{123}$, implying $\theta_3 > 180$ degrees. By Corollary 4.2 for $\Gamma_3$, $3A \leq 3B$.

Now if $\Gamma_3$ leaves $v_{234}$ at or above the horizontal, then Lemma 7.3 gives $2A < 4A$, a contradiction. \hfill \Box

Lemma 7.5. For a (0,0) or (0,1) leaf in a minimizer, as in Figure 14, $1B \leq f(\Gamma_3) \leq 4A$ and $1B < 4A$.

Proof. First, consider case (0,0). Then $2A < 1B$ and $4A < 2B$. Since $v_{123}$ is left of $v_{234}$, $1B < 2B$ and $2A < 4A$. Hence $1B, 4A \in f(\Gamma_2)$.

Second, consider case (0,1). Then $2A < 1B$. Since $v_{123}$ is left of $v_{234}$, $2A < 4A$. Since $(2A, +\infty) \subset f(\Gamma_2)$, again $1B, 4A \in f(\Gamma_2)$.

In both cases, $1B < 3A$ and $3B < 4A$. Corollary 4.4 for $\Gamma_1, \Gamma_2, \Gamma_3$, and for $\Gamma_2, \Gamma_3, \Gamma_4$ gives $1B \leq f(\Gamma_3)$ and $f(\Gamma_3) \leq 4A$, respectively. Since $1B < 3A, 1B < 4A$, as claimed. \hfill \Box

Proposition 7.6. An arc of outer boundary cannot turn vertical downward after leaving a leaf of the smaller region of a minimizer.

Proof. First, consider cases (0,0) and (0,1) of Figure 14. By Lemma 7.5, $1B < 4A$. If $\Gamma_1$ turns vertical downward, $(1B, +\infty) \subset f(\Gamma_1)$. Since by positive pressure $\Gamma_1$ is not a hyperplane, $f(\Gamma_1) \cap f(\Gamma_4) \neq \emptyset$. Corollary 4.4 implies instability, so $\Gamma_1$ cannot turn vertical downward. Similarly, $\Gamma_4$ cannot turn vertical downward.

Second, consider case (0,2) of Figure 15.

If $\Gamma_1$ turns vertical downward, then $(1B, +\infty) \subset f(\Gamma_1)$. Also, $(2A, +\infty) \subset f(\Gamma_2)$, and consideration of $v_{123}$ gives $2A < 1B < 3A$. Therefore, $f(\Gamma_1) \cap f(\Gamma_2) \cap f(\Gamma_3) \neq \emptyset$ (by Lemma 3.1, $\Gamma_3$ is not a vertical line), contrary to Corollary 4.4. Hence $\Gamma_1$ cannot turn vertical downward.

If $\Gamma_4$ turns vertical downward and to the right, then $f(\Gamma_4) = L \cup \{\infty\}$. If $\Gamma_4$ turns vertical downward and to the left, then $(4A, +\infty) \subset f(\Gamma_4)$. Either way, $(2A, +\infty) \subset f(\Gamma_2)$. Consideration of $v_{234}$ gives $4A < 3B$, while $2A < 3B$ since $v_{123}$ is left of $v_{234}$. Also, $3B \neq +\infty$; otherwise $\Gamma_3$ is a concave leftward nodoid or a hyperplane, both disallowed by Lemma 3.1. Thus, $3B \in f(\Gamma_2) \cap f(\Gamma_4)$. Corollary 4.4 for $\Gamma_2, \Gamma_3, \Gamma_4$ implies instability. Hence $\Gamma_4$ cannot turn vertical downward.
Third, consider case \((2, 1)\) of Figure 16. Since \(\infty \in f(\Gamma_2) \cap f(\Gamma_3)\), if either \(\Gamma_1\) or \(\Gamma_4\) goes vertical at all, then there is a separating set involving that arc, \(\Gamma_2\) and \(\Gamma_3\), violating Corollary 4.4.

By Proposition 7.1, every locally stable leaf of larger pressure can be modeled by one of the cases \((0, 0)\), \((0, 1)\), \((0, 2)\) or \((2, 1)\). We conclude that an arc of outer boundary cannot turn vertical downward after leaving any such leaf.

\[\square\]

**Corollary 7.7.** In a minimizer, any arc of outer boundary between two leaves of the smaller region, from a circular cap to such a leaf, or between the two circular caps is graph.

**Proof.** If such an arc goes vertical in its interior, then it is a nodoid or vertical line, by Theorem 3.3. By positive pressure, it must be a nodoid. Therefore it turns vertical downward after leaving either a circular cap or a leaf of the smaller region, contradicting Proposition 6.2 or Proposition 7.6, respectively. \(\square\)

**8. \(1 + k\) double bubbles.**

For any minimal \(1 + k\) double bubble, i.e., an area-minimizing nonstandard double bubble in which one region is connected and the other region has \(k\) components, by Theorem 3.2 the associated tree \(T\) has just one branch from the root and \(k - 1\) leaves above the connected middle region, as in Figure 20.

**Proposition 8.1.** In a \(1 + k\) minimizer, \(k > 1\), in which the larger region is connected, there can be no leaf with left rotation number \(m_1 \geq 3\) or with right rotation number \(m_2 \leq -3\). In particular, with the standard notation of Figure 13, each leaf has rotation pair

\[(m_1, m_2) \in \{(0, 0), (-1, 0), (0, 1), (-2, 0), (0, 2), \pm(1, 2), \pm(2, 1)\}.

\]
Proof. Assume that there is a leaf with \( m_2 \leq -3 \). Let \( m_3 \) measure the rotation of the right endpoint of \( \Gamma_4 \).

Suppose \( \Gamma_4 \) connects to another leaf. If \( m_2 = -3 \) or \(-4\), then by positive pressure \( \Gamma_4 \) is a convex nodoid and \( m_3 \leq -3 \). If \( m_2 \leq -5 \), then by Corollary 7.7 \( m_3 \leq -5 \). In either case, by Proposition 7.1, \( m_4 \) the right rotation number of this adjacent leaf satisfies \( m_4 \leq -3 \) also. By induction, every leaf combinatorially to the right of the original leaf has right rotation number at most \(-3\).

Hence we may assume that \( \Gamma_4 \) connects to the right circular cap. By positive pressure, \( m_3 \leq -2 \), beyond the range of \( \{-1, 0, 1\} \) allowed by Proposition 6.1.

From this contradiction, it follows that \( m_2 \geq -2 \) for each leaf. Symmetrical considerations give \( m_1 \leq 2 \) for each leaf. The second assertion follows from applying these inequalities to the possible \((m_1, m_2)\) pairs of Proposition 7.1. \( \square \)

Define a leaf to be near graph if it has rotation pair \((m_1, m_2)\) \( \in \{(-1, 0), (0, 0), (0, 1)\} \), as in Figure 14.

Proposition 8.2. Consider a 1 + \( k \) minimizer, with the notation of Figure 20, in which the first \( j \), \( 0 \leq j \leq k-1 \), leaves on the left are near graph. If \( m_1 \in \{-1, 0\} \) — necessarily true if \( j > 0 \), or if \( j = 0 \) and \( m_2 \leq 1 \) — then \( m_{2k} \in \{-1, 0\} \) and \( 0B \leq (j+1)B \).

Proof. By Corollary 7.7, \( \Gamma_1, \ldots, \Gamma_k \), the outer boundaries of the middle component indexed from left to right, are graph.

If \( m_2 \leq 1 \) and \( m_1 = 1 \), then \( \Gamma_1 \) turns vertical in its interior, a contradiction. Hence \( m_2 \leq 1 \) — trivially true if \( j > 0 \) — implies \( m_1 \in \{-1, 0\} \), the only remaining possibilities of Proposition 6.1.

Suppose \( m_1 \in \{-1, 0\} \); by Proposition 6.1, \( m_{2k} \in \{-1, 0\} \). If \( m_1 = -1 \), then \( (-\infty, 0B) \subset f(\Gamma_0) \). If \( m_1 = 0 \), then consideration of \( v_{01} \) gives \( 0A < 1A \), implying by Corollary 4.4 for \( \Gamma_0, \Gamma_1 \) that \( 0A \leq 1B \). In either case, Corollary 4.4 for \( \Gamma_0, \Gamma_1 \) gives \( 0B \leq 1B \), the statement for \( j = 0 \).

Now assume \( j > 0 \). Suppose \( 0B \leq iB \), where \( 1 \leq i \leq j \). By Lemma 7.5 with relabelling, \( iB < (i+1)A \). Hence \( 0B < (i+1)A \), implying by Corollary 4.4 for \( \Gamma_0, \Gamma_{i+1} \) that \( 0B \leq (i+1)B \). The statement follows by induction in \( i \). \( \square \)

Corollary 8.3. A 1 + \( k \) minimizer must have at least one leaf above the middle component which is not near graph. In particular, a nonstandard 1 + 1 double bubble cannot be minimizing.

Proof. Suppose all the leaves are near graph (true if \( k = 1 \), when there are no leaves). By Proposition 8.2 applied once to each side, \( m_1 = m_2 = 0 \), and \( 0B \leq kB \). Consideration of \( v_{0k} \) (the right vertex, if \( k = 1 \)) gives \( kB < 0B \), a contradiction. \( \square \)
Lemma 8.4. A $1 + k$ minimizer includes at most one $(2,1)$-modeled leaf: Cases $±(1,2)$ and $±(2,1)$.

Proof. In a $(2,1)$-modeled leaf as in Figure 16, $∞ ∈ f(Γ_2) ∩ f(Γ_3)$. If the minimizer includes more than one such leaf, $\{f^{-1}(∞)\}$ separates $Γ$, violating Corollary 4.4.

Lemma 8.5. In a $1 + k$ minimizer in which the larger region is connected, only a $(0,2)$ leaf may be directly to the left of a $(2,1)$ leaf. Also, every $(0,2)$ leaf in the minimizer must be directly to the left of a $(2,1)$ leaf. The minimizer does not include any $(1,2)$ or $(-2,-1)$ leaves.

Proof. Consider a $(2,1)$ leaf as in Figure 16, and assume that $Γ_1$ connects to another leaf. Let $m_0$ measure the rotation of the (combinatorially) left endpoint of $Γ_1$. By positive pressure, $Γ_1$ is a concave rightward nodoid, so $m_0 ≥ 2$. By Proposition 8.1, the adjacent leaf has rotation pair $(0,2)$ or $(1,2)$, and Lemma 8.4 disallows the latter possibility. Hence indeed, if the minimizer contains a $(2,1)$ leaf, then either that leaf is the leftmost leaf in the minimizer or it is just to the right of a $(0,2)$ leaf.

Now assume there is a leaf with rotation pair $(m_1, m_2) ∈ \{(0,2), (1,2)\}$, as in Figure 15 or Figure 16 with reflection and relabelling. Let $m_3$ measure the rotation of the right endpoint of $Γ_4$.

If $Γ_4$ connects to the right circular cap, by Proposition 6.1 $m_3 ∈ \{-1, 0, 1\}$. Hence $Γ_4$ turns vertical downward after leaving the root, violating Proposition 6.2.

Therefore $Γ_4$ connects to another leaf. By positive pressure, $Γ_4$ is a concave rightward nodoid, implying $m_3 ≤ 2$. For $m_3 ≤ 1$, $Γ_4$ turns vertical downward after leaving the adjacent leaf, violating Proposition 7.6. Thus $m_3 = 2$, whence by Proposition 8.1 the adjacent leaf has rotation pair $(2,1)$.

Therefore, if $(m_1, m_2) = (0,2)$, then the leaf is directly to the left of a $(2,1)$ leaf, as asserted.

If on the other hand $(m_1, m_2) = (1,2)$, then again the leaf is directly to the left of a $(2,1)$ leaf. But now the minimizer includes both a $(1,2)$ and a $(2,1)$ leaf, contradicting Lemma 8.4. Hence the minimizer includes no $(1,2)$ leaves. By symmetry, nor does it include any $(-2,-1)$ leaves. □

Proposition 8.6. A nonstandard double bubble in $\mathbb{R}^n$, $n ≥ 3$, in which the larger region is connected and the smaller region has $k ≥ 1$ components is not minimizing.

Proof. Suppose otherwise and consider the generating curves of the minimizer.

By Corollary 8.3, the minimizer includes at least one leaf which is not near graph. By Proposition 8.1, the possibilities for this leaf, up to horizontal reflection, are $(0,2)$, $(1,2)$ or $(2,1)$. Lemma 8.5 rules out case $(1,2)$ (and,
Figure 21. If the (2,1) leaf is the leftmost leaf, then all other leaves are near graph.

symmetrically, case (−2,−1)). If the minimizer includes a (0,2) leaf, then there is a (2,1) leaf directly to its right, also by Lemma 8.5.

Hence, after horizontal reflection if necessary, the minimizer includes at least one (2,1) leaf. By Lemma 8.4, the minimizer includes exactly one (2,1) leaf, and no (−1,−2) leaves. Therefore, since by Lemma 8.5 every (0,2) leaf is directly to the left of a (2,1) leaf, there is at most one (0,2) leaf, and no (−2,0) leaves. By Lemma 8.5, if there is a leaf to the left of the (2,1) leaf, then it is the (0,2) leaf. If there is no leaf to the left of the (2,1) leaf, then there are no (0,2) leaves.

All leaves not directly to the left of the (2,1) leaf must be near graph, the only possibilities allowed by Proposition 8.1 which still remain.

First, assume the (2,1) leaf is the leftmost leaf, as in Figure 21. By Corollary 7.7, Γ₁ is graph, implying by Proposition 6.1 that \( m₁ = 1 \). Consider the downward normal \( n \) to \( Γ₄ \) at \( v_{234} \). By Lemma 3.1 and Theorem 3.3, \( Γ₂ \) is a concave leftward nodoid, so \( n \) stays above and to the right of \( Γ₂ \). By Corollary 7.4, \( Γ₃ \) leaves \( v_{234} \) below the horizontal, implying that \( n \) is counterclockwise from the downward tangent to \( Γ₁ \) at \( v_{01} \). Since by positive pressure \( Γ₁ \) is a concave rightward nodoid, \( n \) is counterclockwise from every downward tangent to \( Γ₁ \). Therefore, \( n \) stays above \( Γ₁ \) and \( 0A < 4A \). But by Proposition 8.2 applied from the right, \( 4A ≤ 0A \), a contradiction. Hence the (2,1) leaf is not the leftmost leaf.

Second, assume there is a (0,2) leaf to the left of the (2,1) leaf, with the notation of Figure 22. Again, similar arguments using Corollary 7.4 show that the downward normal \( n \) to \( Γ₇ \) at \( v_{567} \) stays to the right of \( Γ₄ \). Since \( v_{123} \) is left of \( v_{234} \), \( n \) is counterclockwise and to the right of the downward normal to \( Γ₁ \) at \( v_{123} \), whence \( 1B < 7A \). But by Proposition 8.2 applied once to each side, \( m₁ = m₂k = 0 \), \( 0B ≤ 1B \) and \( 7A ≤ 0A \). Combining the inequalities yields \( 0B < 0A \), a clear impossibility when \( m₁ = m₂k = 0 \). Hence there can be no leaf to the left of the (2,1) leaf.

Therefore, a \( 1+k \) minimizer cannot include a (2,1) leaf, a contradiction. Thus indeed, a \( 1+k \) bubble in which the larger region is connected cannot be minimizing. \( \Box \)

Theorem 9.1 (Double Bubble Conjecture in \( \mathbb{R}^4 \)). In \( \mathbb{R}^4 \), the standard double bubble is the unique area-minimizing double bubble.

Proof. For equal volumes, both regions are connected by Proposition 2.5. By Corollary 8.3, the area-minimizing double bubble is the standard double bubble.

For unequal volumes, the larger region is connected by Proposition 2.5, and the smaller region has a finite number of components by Corollary 2.4. By Proposition 8.6, the area-minimizing double bubble is the standard double bubble. \( \Box \)

Theorem 9.2 (Double Bubble Conjecture for disparate volumes). In \( \mathbb{R}^n \), \( n \geq 3 \), the standard double bubble is the unique area-minimizing double bubble for prescribed volumes \( v, w \), with \( v > 2w \).

Proof. The larger region is connected by Corollary 2.2, and the smaller region has a finite number of components by Corollary 2.4. By Proposition 8.6, an area-minimizing double bubble must be the standard double bubble. \( \Box \)

Acknowledgements. The authors were the members of the 1999 Geometry Group of the Williams College SMALL undergraduate research project. Their work was partially funded by the National Science Foundation. They gratefully thank Professor Frank Morgan for his very helpful comments and advice.

References

B. Reichardt, C. Heilmann, Y. Lai, and A. Spielman


Received February 29, 2000 and revised November 28, 2000.

EECS DEPARTMENT
COMPUTER SCIENCE DIVISION
UC BERKELEY
E-mail address: breic@cs.berkeley.edu

DEPARTMENT OF STATISTICS
IOWA STATE
E-mail address: heilmann@iastate.edu

DEPARTMENT OF MATHEMATICS
UC DAVIS
E-mail address: lai@math.ucdavis.edu

DEPARTMENT OF MATHEMATICS
WILLIAMS COLLEGE
WILLIAMSTOWN, MA 01267
E-mail address: aspielma@wso.williams.edu
A TOPOLOGICAL APPROACH TO INVERSE AND REGULAR SEMIGROUPS

Benjamin Steinberg

Work of Ehresmann and Schein shows that an inverse semigroup can be viewed as a groupoid with an order structure; this approach was generalized by Nambooripad to apply to arbitrary regular semigroups. This paper introduces the notion of an ordered 2-complex and shows how to represent any ordered groupoid as the fundamental groupoid of an ordered 2-complex. This approach then allows us to construct a standard 2-complex for an inverse semigroup presentation.

Our primary applications are to calculating the maximal subgroups of an inverse semigroup which, under our topological approach, turn out to be the fundamental groups of the various connected components of the standard 2-complex. Our main results generalize results of Haatja, Margolis, and Meakin giving a graph of groups decomposition for the maximal subgroups of certain regular semigroup amalgams. We also generalize a theorem of Hall by showing the strong embeddability of certain regular semigroup amalgams as well as structural results of Nambooripad and Pastijn on such amalgams.

1. Introduction.

In the fifties, there were two attempts to axiomatize the underlying structure of pseudogroups of diffeomorphisms of manifolds. One approach, by Wagner (and independently by Preston [16]), was via inverse semigroups; the other, by Ehresmann, was via ordered groupoids, namely the so-called inductive groupoids popularized amongst semigroup theorists by Schein [18]. It is fascinating that some results are proved more easily via the inverse semigroup approach, while others are more naturally proved from the point of view of ordered groupoids. In his seminal paper [11], Nambooripad extended this equivalence to an equivalence between regular semigroups and, what we shall call in this paper, $r$-inductive groupoids. Recent work [15] emphasizes the importance of inverse semigroups in the theory of $C^*$-algebras.

In this paper, building on the idea that any groupoid can be realized as the fundamental groupoid of a 2-complex, we introduce the notion of an ordered 2-complex and show that any ordered groupoid can be realized as
the fundamental ordered groupoid of an ordered 2-complex. In the process, we introduce the notion of a presentation of an ordered groupoid. While an inverse semigroup presentation is different from an ordered groupoid presentation, we show how to construct an ordered groupoid presentation from any inverse semigroup presentation by constructing what we call the standard ordered 2-complex of an inverse semigroup presentation (generalizing the usual notion from group presentations). This complex is closely related to the Munn and right letter mapping representations of the inverse semigroup.

The maximal subgroups of the inverse (regular) semigroup will turn out to be the fundamental groups of the connected components of an ordered 2-complex representing the semigroup. This will allow us to obtain a quick topological proof that if a maximal subgroup of a finitely presented inverse semigroup acts on the left of its $R$-class with finite quotient, then it is finitely presented.

We also introduce the Schützenberger complex of an inverse semigroup presentation. This complex is a $\pi_1$-trivial covering space of the standard ordered 2-complex of the presentation and plays a role in the theory similar to that played by the Cayley complex (which it generalizes) in group theory. The 1-skeleton of the Schützenberger complex is the union of all the Schützenberger graphs of the presentation [24]. This complex and the standard ordered 2-complex, both of which are built from semigroup theoretic tools, provide a link between the semigroup and ordered groupoid approaches to the subject.

Our main result is a structure theorem for maximal subgroups of certain amalgams. In [3], it is proved that a maximal subgroup of a full amalgam of regular semigroups has a certain graph of groups decomposition. In this paper, we are able to weaken the restrictions and, at the same time, obtain a much simpler proof. The proof of [3] relies on Bass-Serre theory whence they have to construct an action of the maximal subgroup on a tree with appropriate stabilizers; they use normal forms for the amalgamated product in order to do this. Our approach is to show that the amalgamated product can be represented by a topological space which is a segment of disconnected 2-complexes; the structure of the maximal subgroups then follows from an obvious decomposition of such a graph of 2-complexes into a graph of connected 2-complexes. In the process, we extend a result of Hall [4] on the strong embeddability of regular semigroup amalgams as well as structural results of Nambooripad and Pastijn [12] on such. We obtain further results for certain inverse semigroup amalgams, overlapping at times with results of Bennet [1]. We also obtain results for certain amalgams in the category of inverse semigroups and prehomomorphisms. For some related results on inverse semigroup amalgams, see [22].
A TOPOLOGICAL APPROACH TO INVERSE 369

Future work, will show that the various structure theorems for idempotent pure morphisms of inverse semigroups correspond to facts about when a morphism of ordered 2-complexes can be extended nicely to a covering.

2. Preliminaries.

If $A$ is a set, we let $\tilde{A} = A \cup A^{-1}$. Then $\tilde{A}^+$ will denote the free semigroup and $\tilde{A}^*$ the free monoid with involution on $A$.

If $S$ is a semigroup, an element $e \in S$ is an idempotent if $e^2 = e$. The set of idempotents of $S$ will be denoted $E(S)$. An element $s \in S$ is called regular if $s = sts$ for some $t \in S$. A semigroup in which each element is regular is called a regular semigroup. If $s$ and $t$ are such that $sts = s$ and $tst = t$, then $s$ and $t$ are said to be inverses of each other. One can show that an element is regular if and only if it has an inverse [6]. A semigroup $S$ is called an inverse semigroup if each element $s \in S$ has a unique inverse (denoted $s^{-1}$) or, equivalently, if it is regular and has commuting idempotents [7]. The notation $S^1$ is used for the monoid obtained by adding an identity to $S$. We refer the reader to [7] for more on inverse semigroups. In this paper, we shall primarily focus on inverse semigroups, only occasionally speaking of more general regular semigroups.

We briefly recall Green’s relations [6]. If $S$ is a semigroup, then we write $s \leq_R t$ if $sS^1 \subseteq tS^1$; this is a preorder and we use $R$ for the associated equivalence relation. The preorder $\leq_L$ and equivalence relation $L$ are defined dually. One sets $H = R \cap L$ and $D = R \circ L$; $D$ is the smallest equivalence relation containing $R$ and $L$. If $e \in E(S)$, then the $H$-class of $e$ is the maximal subgroup with identity $e$ [6]. For an inverse semigroup, $s \mathrel{R} t$ if and only if $ss^{-1} = tt^{-1}$ [7] and dually for $L$.

If $I$ is an inverse semigroup, there is a natural partial order on $I$ defined by $s \leq t$ if $s = ss^{-1}t$. This order is well-known to be compatible with the operations of multiplication and taking inverses [7].

Inverse semigroups can be viewed as a variety of unary semigroups and as such there is a free inverse semigroup on any set as well as a notion of inverse semigroup presentations; see [24] for some graph-theoretical techniques for working with inverse semigroup presentation.

A morphism $\varphi$ of partially ordered sets is called an order embedding if $\varphi(x) \leq \varphi(y)$ if and only if $x \leq y$. A surjective order embedding is called an order isomorphism.

3. Ordered graphs.

Following the Serre convention [19], we define a graph $X$ to consist of: A set $V(X)$ of vertices; a set $E(X)$ of edges; an involution $e \mapsto e^{-1}$ on $E(X)$; and a function $d : E(X) \to V(X)$ which selects the initial vertex of an edge (the terminology $d$ is chosen to suggest the word domain). We define
\(r : E(X) \to V(X)\) by \(r(e) = d(e^{-1})\) (and \(r(e)\) is called the \textit{terminal} vertex of \(e\)). One defines, for \(v \in V(X)\), \(\text{Star}(v) = d^{-1}(v)\). In general, we shall require \(e^{-1} \neq e\), the exception being when we view groupoids as graphs in which case local group elements of order 2 will have this property. Our reason for choosing this unorthodox notation for graphs is because many of our graphs will, in fact, be groupoids, where the terms domain and range have obvious meanings.

Graph morphisms are defined in the obvious way. If \(\varphi : X \to Y\) is a graph morphism and, for each \(v \in V(X)\), \(\varphi : \text{Star}(v) \to \text{Star}(\varphi(v))\) is, respectively, injective, surjective, bijective, then \(\varphi\) is called, respectively, an \textit{immersion}, a \textit{fibration}, a \textit{covering}.

An \textit{ordered graph} is a graph \(X\) such that: \(V(X)\) and \(E(X)\) are partially ordered sets; the involution and \(d\) preserve order (whence \(r\) preserves order); and if \(e \in E(X)\), \(v \in V(X)\) are such that \(v \leq d(e)\), then there is a unique edge of \(X\), denoted \((v|e)\), with \(d(v|e) = v\) and \((v|e) \leq e\), called the \textit{restriction} of \(e\) to \(v\). Note that this third condition implies that the edges of \(\text{Star}(v)\) are incomparable for any vertex \(v\). Observe that if \(v \leq d(e)\), then \((v|e)^{-1} = (r(v|e)|e^{-1})\). Morphisms of ordered graphs are graph morphisms which preserve the partial order.

If \(v_1, v_2 \in V(X)\) and \(v_1 \leq v_2\), we can define a \textit{restriction map} \(\text{res}_{v_2}^{v_1} : \text{Star}(v_2) \to \text{Star}(v_1)\) by \(\text{res}_{v_2}^{v_1}(e) = (v_1|e)\). The above, together with straightforward reasoning, shows that these restriction maps satisfy the following three properties:

\[
\begin{align*}
(1) & \quad v_1 \leq v_2 \leq v_3 \implies \text{res}_{v_2}^{v_1}\text{res}_{v_3}^{v_2} = \text{res}_{v_3}^{v_1}; \\
(2) & \quad r(\text{res}_{v_2}^{v_1}(e)) \leq r(e); \\
(3) & \quad \text{res}_{v_2}^{v_1}(e)^{-1} = r(\text{res}_{v_2}^{v_1}(e))(e^{-1}).
\end{align*}
\]

Any morphism of ordered graphs automatically preserves these restrictions maps.

Conversely, given a graph \(X\) such that \(V(X)\) is partially ordered and given maps \(\text{res}_{v_2}^{v_1} : \text{Star}(v_2) \to \text{Star}(v_1)\), whenever \(v_1 \leq v_2\), satisfying the above three properties, we can turn \(X\) into an ordered graph by defining, for \(e_1, e_2 \in E(X)\), \(e_1 \leq e_2\) if \(d(e_1) \leq d(e_2)\) and \(e_1 = \text{res}_{d(e_2)}^{d(e_1)}(e_2)\). We shall use freely throughout these two equivalent formulations of the definition of an ordered graph.

An \textit{ordered subgraph} of an ordered graph is a subgraph which is an ordered graph with the induced ordering (that is, a subgraph closed under restrictions).

\section*{4. Ordered groupoids.}

If \(G\) is an ordered graph, then two edges \(e_1, e_2\) are said to be \textit{composable} if \(r(e_1) = d(e_2)\). An \textit{ordered groupoid} is an ordered graph with an associative
A TOPOLOGICAL APPROACH TO INVERSE 371

multiplication on composable edges satisfying, whenever the compositions make sense:

(1) $xx^{-1}y = y$;
(2) $x_1 \leq y_1, x_2 \leq y_2 \implies x_1x_2 \leq y_1y_2$.

The first condition implies that at each vertex $v$ there is a unique identity which will be denoted $1_v$. It is easy to show that the map $v \mapsto 1_v$ is an order isomorphism.

In terms of the restriction maps, the second condition becomes: If $x_1x_2$ is defined and $v \leq d((x_1)$, then

$$(v|x_1x_2) = (v|x_1)(r(v|x_1)|x_2).$$

Remember that, for groupoids, we do allow $e = e^{-1}$.

An ordered subgroupoid of an ordered groupoid is a subgroupoid which is an ordered subgraph.

An inductive groupoid is an ordered groupoid $G$ in which $V(G)$ is a semi-lattice; we shall call $G$ r-inductive if $V(G)$ is a biordered set [11]. These definitions make perfect sense for ordered graphs and we shall extend their domains to that realm. The notions of ordered and inductive groupoids are due to Ehresmann [2]; see also Schein [18].

Any inverse semigroup $I$ can be realized as an inductive groupoid as follows: $V(I) = E(I)$; $E(I) = I$; the order is the natural partial order; the inverse is the inverse; $d(s) = ss^{-1}$; and the product is the usual multiplication. Conversely, given an inductive groupoid $I$, we can define an inverse semigroup whose elements are $E(I)$. If $s,t \in E(I)$, we define

$$st = (r(s) \wedge d(t)|s^{-1})^{-1}(r(s) \wedge d(t)|t).$$

These two constructions are inverse to each other; see [7] for more details. We observe that the $\mathcal{R}$-class of an idempotent $e \in E(I)$ is $\text{Star}(e)$, the $\mathcal{H}$-class of $e$ is the local group at $e$ (see below for the definition), and the $\mathcal{D}$-class of $e$ is the connected component of the groupoid containing $e$. Inverse semigroup homomorphisms correspond to ordered groupoid morphisms which preserve the inductive structure.

There is a similar, but more complicated, correspondence between regular semigroups and r-inductive groupoids [11].

A groupoid in the usual sense is an ordered groupoid under the equality ordering. In particular, any group is an inductive groupoid.

If $v \in V(G)$ is a vertex of an ordered groupoid $G$, then

$$G_v = \{g \in E(G)|d(g) = v = r(g)\}$$

is a group called the local group or maximal subgroup of $G$ at $v$.

An order ideal in a partially ordered set is a subset $I$ such that $x \leq y \in I$ implies $x \in I$. If $G$ is an ordered groupoid, we use $\text{Id}(G)$ to denote
the ordered subgroupoid consisting of all vertices and identities. It can be shown [7] that \( E(\text{Id}(G)) \) is an order ideal in \( E(G) \).

5. Paths and free ordered groupoids.

A path in a graph is defined in the usual way: we allow an empty path at each vertex. If \( p \) is a path, we shall use \( d(p) \) for its initial vertex, \( r(p) \) for its terminal vertex, and \( p^{-1} \) for the reverse path. A path \( p \) is said to be closed if \( d(p) = r(p) \). If \( p \) is a closed path, we say that a path \( p' \) is a cyclic conjugate of \( p \) if \( p = st \) and \( p' = ts \) (that is, \( p \) and \( p' \) are “the same path” starting from different vertices). A closed path is called simple if the only repetition of vertices when tracing the path occurs when one reaches the end. Connected components of a graph are defined in the usual way. Using terminology suggestive of semigroup theory, connected components will also be called \( \mathcal{D} \)-classes. If \( v \) is a vertex, then \( \mathcal{D}_v \) will denote the \( \mathcal{D} \)-class of \( v \). Note that all maximal subgroups at vertices of the same \( \mathcal{D} \)-class are isomorphic.

We say that a path \( p \) is obtained from a path \( p' \) by an elementary homotopy if \( p \) is gotten from \( p' \) by inserting or deleting a subpath of the form \( ee^{-1} \). One says that paths \( p \) and \( p' \) are homotopic if \( p \) can be turned into \( p' \) by a finite sequence of elementary homotopies; this is an equivalence relation and equivalent elements are coterminal. It is well-known that this definition is equivalent to asking that the paths be homotopic (in the topological sense) in the geometric realization of the graph.

We now aim to prove that the fundamental groupoid of an ordered graph is an ordered groupoid in a natural way. We shall need several lemmas to achieve this. First we define restriction maps for paths which are compatible with homotopy. If \( X \) is an ordered graph, \( p = e_1 e_2 \ldots e_n \) is a path, and \( v \leq d(p) \) we define \( (v|p) \) as follows: First set \( e'_1 = (v|e_1) \); then, inductively, set \( e'_{i+1} = (r(e'_i)|e_{i+1}) \) for \( 1 \leq i \leq n-1 \). It is straightforward to check that

\[(v|p) = \text{res}^{v}_{d(p)}(p) = e'_1 \ldots e'_n\]

is a path starting at \( v \). Observe that \( r(v|p) = r(e'_n) \leq r(e_n) \). One can show by induction that if \( v_1 \leq v_2 \leq v_3 \), then \( \text{res}^{v_1}_{v_2} \text{res}^{v_2}_{v_3} = \text{res}^{v_1}_{v_3} \).

**Lemma 5.1.** Suppose \( p = e_1 \ldots e_n \) and \( p' = f_1 \ldots f_m \) are paths in an ordered graph with \( r(p) = d(p') \) and \( v \leq d(p) \). Then:

1. \((v|pp') = (v|p)(r(v|p)|p')\)
2. \(\text{res}^{r(v|p)}_{r(p)}(p^{-1}) = (v|p)^{-1}\).

*Proof.* The inductive nature of the definition of the restriction of a path shows that to prove 1, it suffices to show that the \((n+1)\)st edge of \((v|pp')\) is \((r(v|p)|f_1)\). Letting \( e'_1, \ldots, e'_n \) be as above, we see that the \((n+1)\)st edge is \((r(e'_n)|f_1) = (r(v|p)|f_1)\) as desired.
We note that (2) follows immediately from (1) using induction (the base case being part of the definition of an ordered graph).

We now see how elementary homotopies behave under restriction.

**Lemma 5.2.** Suppose that e is an edge of an ordered graph and \( v \leq d(e) \). Then \( (v|ee^{-1}) = (v|e)(v|e)^{-1} \).

*Proof.* By definition, \( (v|ee^{-1}) = (v|e)(v|e) \), but the right-hand side is \( (v|e)(v|e)^{-1} \). □

**Corollary 5.3.** If \( p, p' \) are homotopic paths in an ordered graph and \( v \leq d(p) = d(p') \), then \( (v|p) \) and \( (v|p') \) are homotopic.

*Proof.* Follows immediately from Lemmas 5.1 and 5.2. □

What we have now proved is that if \( X \) is an ordered graph, then the fundamental groupoid of \( X \) is an ordered groupoid. That is, we define an ordered groupoid \( \Pi_1(X) \) by taking \( V(\Pi_1(X)) = V(X) \) and \( E(\Pi_1(X)) \) to be the set of all homotopy classes of paths in \( X \); \( d \) and the involution are defined in the obvious way. We can then use path composition and the path restriction maps (defined above) to turn \( \Pi_1(X) \) into an ordered groupoid, Lemmas 5.1 and 5.2, together with Corollary 5.3, providing the necessary verifications. This groupoid is called the fundamental ordered groupoid of \( X \).

Its underlying groupoid is just the usual (simplicial) fundamental groupoid of \( X \) (it is the subgroupoid of the topological fundamental groupoid whose vertices are vertices of \( X \) and whose edges are homotopy classes of edge paths). Note that \( X \) embeds in \( \Pi_1(X) \) in a natural way and that if \( X \) is an \((r)\)-inductive graph, then so is \( \Pi_1(X) \).

We leave the following ordered analog of a standard topological fact to the reader:

**Proposition 5.4.** Let \( X \) be an ordered graph, \( G \) an ordered groupoid, and \( \rho : X \to G \) a morphism of ordered graphs. Then there is a unique ordered groupoid morphism \( \overline{\rho} : \Pi_1(X) \to G \) extending \( \rho \). Furthermore, \( \rho \) is an immersion, fibration, covering if and only if \( \overline{\rho} \) is. In the \((r)\)-inductive setting, \( \rho \) preserves the \((r)\)-inductive structure if and only if \( \overline{\rho} \) does.

Proposition 5.4 shows that \( \Pi_1(X) \) is the free ordered groupoid generated by \( X \). To make this precise, we say that an ordered graph \( X \) generates an ordered groupoid \( G \) if there is a morphism \( \rho : \Pi_1(X) \to G \) which is an order isomorphism on vertex sets and surjective on edge sets. We warn the reader that in the \((r)\)-inductive setting, this concept does not correspond to a choice of semigroup generators. The kernel

\[
\ker \rho = \{ p \in \Pi_1(X) \mid \rho(p) = 1_{d(p)} \}
\]
is then an order ideal. This leads us to the notion of an ordered groupoid presentation.

If \( X \) is an ordered graph and \( R \) is a collection of closed paths in \( X \), then \( R \) is called a set of \textit{relators} if \( p \in R \) and \( v \leq d(p) \) implies \( (v|p) \in R \) (that is, \( R \) is \textit{closed under restrictions}). It then follows that the image of \( R \) in \( \Pi_1(X) \) is an order ideal. The ordered groupoid presented by generators \( X \) and relators \( R \), denoted \( \langle X | R \rangle \), is then the quotient of \( \Pi_1(X) \) obtained by identifying two paths \( p, p' \) if \( p \) can be gotten from \( p' \) by a finite sequence of elementary homotopies and insertions or deletions of paths in \( R \cup R^{-1} \) (or, equivalently, cyclic conjugates of such). The fact that \( R \) is closed under restrictions allows one to show, similarly to the argument with elementary homotopies, that the path restrictions induce well-defined maps on the quotient groupoid.

Note that every ordered groupoid \( G \) has a presentation since we can choose \( G \), itself, as the generating graph and all paths mapping into the kernel of the canonical surjection as the set of relators. We call this the \textit{multiplication table presentation}.

If \( G \) is a group with generators \( A \) and relators \( R \), then we can view \( \tilde{A} \) as a bouquet of circles labeled by \( \tilde{A} \), that is to say, as a graph \( X \) with a single vertex and edge set \( \tilde{A} \); we impose the equality ordering. Then \( \Pi_1(X) \) is the free group on \( A \) and \( G \) is the ordered groupoid presented by \( A \) and \( R \).

We remark that any ordered groupoid generated by an \((r-)\)-inductive graph is \((r-)\)-inductive.


In the previous section, we saw how to add relations to the fundamental ordered groupoid of an ordered graph; in this section, we look at the geometric analog: Gluing in 2-cells. There are several ways in which we could do this. For now, we shall choose a rather general method; future developments, especially the consideration of higher dimensional complexes, might make it necessary to impose more stringent restrictions.

From now on, we shall be a little less formal about topological ideas (thus requiring a greater familiarity with such on the part of the reader) to avoid the cumbersome task of defining coverings of 2-complexes combinatorially.

We define a 2-cell to be a regular \( n \)-gon in \( \mathbb{R}^2 \), \( n > 0 \), where, by convention, we take a regular 1-gon to be the unit disk with a single vertex \( (0,1) \) and one edge pair \( \{e, e^{-1}\} \), and we take a regular 2-gon to be the unit disk with vertices \( (-1,0) \) and \( (1,0) \) with the obvious two edge pairs. If \( c \) is a 2-cell, its boundary will be denoted \( \partial c \).

An \textit{ordered 2-complex} \( X \) consists of an ordered graph \( X^{(1)} \) (the \textit{1-skeleton}) and a set \( C(X) \) consisting of pairs \((c, f_c)\) where \( c \) is a 2-cell and \( f_c : \partial c \to X^{(1)} \) is a graph morphism called the \textit{attaching map} for \( c \). Topologically, \( c \) is glued to \( X^{(1)} \) via \( f_c \). We shall write \( V(X) \) and \( E(X) \) for the vertex and
edge sets of \( X^{(1)} \). A \textit{defining path} for \((c, f_c) \in C(X)\) is the image under \( f_c \) of a simple, nonempty, closed path in \( \partial c \). We then require, for any defining path \( p \) for \((c, f_c) \) and \( w \leq d(p) \), that \((w|p)\) is a defining path for some \((c', f_{c'}) \in C(X)\). Future work may show that we actually want to order \( C(X) \) and have a restriction map.

Given an ordered 2-complex, we can define a set of relators by

\[ R_X = \{ p \mid p \text{ is a defining path for some } (c, f_c) \in C(X) \}; \]

the definition of an ordered 2-complex insures that \( R_X \) is closed under restrictions. We define

\[ \Pi_1(X) = \langle X|R_X \rangle \]

to be the \textit{fundamental ordered groupoid} of \( X \). As before, one can verify that two paths in \( X^{(1)} \) are homotopic topologically in (the geometric realization of) \( X \) if and only if they are equivalent in \( \Pi_1(X) \); that is, \( \Pi_1(X) \) embeds naturally as a subgroupoid of the topological fundamental groupoid. Once again, we have the standard (simplicial) fundamental groupoid if we ignore the ordering.

If \( v \in V(X) \), then \( \pi_1(X, v) \) will be used to denote the local group at \( v \) and is called the \textit{fundamental group} of \( X \) at \( v \). If \( Y \) is a connected 2-complex, then \( \pi_1(Y) \) will be used to denote the abstract group to which all fundamental groups of \( Y \) are isomorphic. The majority of 2-complexes which we shall consider will not be connected and so we cannot in general speak of the fundamental group of \( X \). An ordered 2-complex will be called \textit{\( \pi_1 \)-trivial} if its fundamental group at each vertex is trivial (or, equivalently, each connected component is simply connected).

If \( G \) is a group given by generators and relators, then its standard 2-complex with the equality ordering has fundamental ordered groupoid exactly the group. More generally, it is immediate how to realize any ordered groupoid given by generators and relators as the fundamental ordered groupoid of an ordered 2-complex.

We shall call an ordered 2-complex \( (r-) \)-inductive if its 1-skeleton is such. In this case, the fundamental groupoid will be \( (r-) \)-inductive as well. It follows that any inverse (regular) semigroup can be represented as the fundamental ordered groupoid of an ordered 2-complex, generalizing the usual result for groups.

Morphisms of ordered 2-complexes are defined in the obvious way. Let \( \varphi : X \rightarrow Y \) be a morphism of ordered 2-complexes. Then there is an induced morphism \( \varphi_* : \Pi_1(X) \rightarrow \Pi_1(Y) \). If \( \varphi \) is a covering morphism in the topological sense, then \( \varphi_* \) will be a covering morphism, but the converse fails. In general, \( \varphi_* \) being a, respectively, immersion, fibration, covering is equivalent to \( \varphi \) having, respectively, at most one lift to any vertex of each path in \( Y \), unique path-lifting, path-lifting.
7. The Schützenberger complex.

In this section, we construct a \( \pi_1 \)-trivial covering of an ordered 2-complex which plays a role in the theory analogous to that of the Cayley complex (or universal covering) in group theory. The goal of this section is to prove the following theorem:

**Theorem 7.1.** Let \( X \) be an ordered 2-complex. Then there is a \( \pi_1 \)-trivial ordered 2-complex \( \tilde{X} \) and a surjective ordered covering \( \varphi : \tilde{X} \to X \). Moreover, for each vertex \( v \in V(X) \), there is a connected component \( \tilde{X}_v \) and a free action of \( \pi_1(X,v) \) on \( \tilde{X}_v \) such that \( D_v = \pi_1(X,v)/\tilde{X}_v \).

Let \( X \) be an ordered 2-complex. Then the Schützenberger graph of \( X \), denoted \( \tilde{X}(1) \), is defined as follows:

1. \( V(\tilde{X}(1)) = E(\Pi_1(X)) \);
2. \( E(\tilde{X}(1)) = \{ (g,x) \mid g \in E(\Pi_1(X)), x \in E(X), \text{ and } r(g) = d(x) \} \).

We define \( d(g,x) = g \) and \( (g,x)^{-1} = (gx,x^{-1}) \). The order on \( V(\tilde{X}(1)) \) is as on \( E(\Pi_1(X)) \), while the order on \( E(\tilde{X}(1)) \) is the product order. It is straightforward to check that this is an ordered graph.

For the case of the standard ordered 2-complex of an inverse semigroup presentation (to be defined shortly), the connected components of \( \tilde{X}(1) \) will be the usual Schützenberger graphs of the various \( R \)-classes whence the name. In the case that \( X \) is the standard 2-complex of a group presentation, \( \tilde{X}(1) \) is the usual Cayley graph.

**Lemma 7.2.** The ordered graph morphism \( \varphi : \tilde{X}(1) \to X(1) \) given on vertices by \( g \mapsto r(g) \) and on edges by projection to the second coordinate is a surjective covering.

**Proof.** It is clear that the map is an ordered graph morphism. To see it is a cover, if \( v \in V(X) \), \( g \in V(\tilde{X}(1)) \) with \( r(g) = v \), and \( x \in E(X) \) with \( d(x) = v = r(g) \), then \( (g,x) \) is the unique lift of \( x \) to \( g \). It follows \( \varphi \) is a covering. Surjectivity is clear. \( \square \)

We now define the Schützenberger complex to be the ordered 2-complex with 1-skeleton \( \tilde{X}(1) \) and whose collection of 2-cells consists of all possible lifts of 2-cells of \( X \) under the 1-skeleton covering \( \varphi \). It is clear \( \tilde{X} \) is an ordered 2-complex and \( \varphi \) induces a topological covering. If \( X \) is the standard 2-complex of a group presentation, the resulting complex is the Cayley complex of the presentation.

We observe that the connected components of \( \tilde{X} \) correspond to the vertices of \( X \); that is, for each \( v \in V(X) \), there is a connected component \( \tilde{X}_v \) of \( \tilde{X} \) whose vertices consists of \( \text{Star}(v) \), and all connected components are of this form.
If \( v, v' \in V(\pi) \) with \( v \leq v' \), one can check that the restrictions induce an ordered graph morphism from \( \tilde{X}^{(1)}_{x} \) to \( \tilde{X}^{(1)}_{v} \) which can be extended to an ordered 2-complex morphism in a natural way. While this will not be important in our current work, we do mention that it is connected with the notion of an ordered forest as per [25].

We now show that \( \tilde{X} \) is \( \pi_{1} \)-trivial. Indeed, a path in \( \tilde{X} \) is of the form

\[
(g, x_{1})(gx_{1}, x_{2}) \cdots (gx_{1} \cdots x_{n-1}, x_{n})
\]

where \( g \in \Pi_{1}(\pi) \) with \( d(g) = v \) and \( p = x_{1} \cdots x_{n} \) is a path in \( \pi \) with \( d(p) = r(g) \). To be closed, we must have \( gp = g \) in \( \Pi_{1}(\pi) \) whence \( p \) is a null homotopic, closed path in \( \pi \). The homotopy killing \( p \) can then be lifted to \( \tilde{X} \); it follows \( \tilde{X} \) is \( \pi_{1} \)-trivial.

Our next goal is to show that \( \Pi_{1}(\pi) \) acts freely on \( \tilde{X} \) and that the quotient can be identified with \( \pi \). In the non-ordered case this is standard, so we don’t go into great detail.

Let \( G \) be an ordered groupoid and \( X \) an ordered 2-complex. Then a left action (analogous to the case of an ordered groupoid acting on an ordered graph morphism \([21]\) (\( \pi, A \)) of \( G \) on \( X \) consists of the following data: First we require an ordered graph morphism \( \pi : X^{(1)} \rightarrow \text{Id}(G) \). Now define an ordered 2-complex \((G, X)\) by:

\[
V(G, X) = \{(g, v) \mid g \in E(G), v \in V(X), \text{ and } r(g) = \pi(v)\};
\]

\[
E(G, X) = \{(g, x) \mid g \in E(G), x \in E(X), \text{ and } r(g) = \pi(x)\}.
\]

We let \((g, x)d = (g, xd)\) and \((g, x)^{-1} = (g, x^{-1})\); the order is the product order. The 2-cells are of the form \((g, c)\) where \( c \) is a 2-cell of \( X \) whose attaching map has image in \( \pi^{-1}(r(g)) \). It is straightforward to check that this is an ordered 2-complex. We then require an ordered 2-complex morphism \( A : (G, X) \rightarrow (\pi) \) (which we normally denote by left multiplication, \( A(g, x) = gx \)) such that the following axioms hold (we use \( \exists gx \) if \( r(g) = \pi(x) \)):

\[\text{A1} \quad \text{If } \exists gx, \text{ then } \pi(gx) = d(g); \]
\[\text{A2} \quad \text{if } \exists g_{1}g_{2}, \exists gx, \text{ then } g_{1}(g_{2}x) = (g_{1}g_{2})x; \]
\[\text{A3} \quad 1_{x} = x, v \in V(G), \text{ whenever it is defined}; \]

where, in the above, \( g, g_{1}, g_{2} \in E(G) \) and \( x \) is an \( n \)-cell of \( X \), \( n = 0, 1, 2 \). The action is said to be free if \( gx = x \) implies \( g \) is an identity for such an \( x \).

We now define an action of \( \Pi_{1}(\pi) \) on \( \tilde{X} \) by defining \( \pi : \tilde{X} \rightarrow X \) to be the map which takes \( \tilde{X} \) to \( v \). The map \( A \) is defined by letting, for \( g, g' \in E(\Pi_{1}(\pi)) \) with \( r(g) = d(g') \), \( A(g, g') = gg' \). For \( g, g' \in E(\Pi_{1}(\pi)) \), \( x \in X \) with \( r(g) = d(g') \) and \( r(g') = d(x) \), we define \( g', x = (gg', x) \). The map is extended to 2-cells in the obvious way. This action is clearly free since if \( gg' = g' \), then \( g \) is an identity. We leave it to the reader to check the details.
Observe that if \( v \in v(X) \), then \( \pi_1(X,v) \) acts freely on \( \tilde{X}_v \). Hence, if this action is co-compact (has compact quotient), \( \pi_1(X,v) \) is finitely presented.

The following proposition is a standard fact about the action of the fundamental groupoid on the universal cover:

**Proposition 7.3.** Let \( \varphi : \tilde{X} \to X \) be the covering projection. Then two \( n \)-cells are identified by \( \varphi \) if and only if they are in the same orbit under the action of \( \Pi_1(X) \). In fact, if \( v \in V(X) \), then \( D_v = \pi_1(X,v)/\tilde{X}_v \).

One can even do better by showing that if we define an order on the quotient complex \( \Pi_1(X)/\tilde{X} \) by \( [x] \leq [x'] \), with \( x \in V(X) \) or \( x \in E(X) \), if, for each \( y' \in [x'] \), there exists \( y \in [x] \) such that \( y \leq y' \) (where we use brackets for orbits), then this order turns the topological homeomorphism of \( X \) and \( \Pi_1(X)/\tilde{X} \) into an isomorphism of ordered 2-complexes. As we shall not use this fact in the sequel, we leave its verification as a tedious exercise.

8. **The standard ordered 2-complex of an inverse semigroup presentation.**

Let \( I = \text{Inv} \langle A|R \rangle \) be an inverse semigroup presentation. We now construct an ordered 2-complex \( X \) which we call the *standard ordered 2-complex* of the presentation \( \text{Inv} \langle A|R \rangle \) of \( I \). The vertices of \( X \) are \( \mathcal{E}(I) \); this makes it clear that our complex will not necessarily be finite for a finite presentation and might not be, in any sense, effectively constructible. The edges are pairs \((e,x)\) where \( e \in \mathcal{E}(I) \), \( x \in \tilde{A} \) and \( ex \in R e \). We define \( d(e,x) = e \) and \((e,x)^{-1} = ((ex)^{-1}ex, x^{-1}) \). We call \( x \) the *label* of \((e,x)\) and extend the notion to paths in the obvious way. If \( I \) is a group, then one has a wedge of circles labeled by the elements of \( A \) (and their inverses). We define a partial order by \( (e,x) \leq (f,y) \) if \( e \leq f \) and \( x = y \). It is straightforward to see that \( X^{(1)} \) satisfies the axioms of an inductive graph. Note that the inverse and restriction maps respect labels. If \( A \) is a finite set, then \( X^{(1)} \) is locally finite.

**Proposition 8.1.** Let \( e \in V(X) \) and \( w \in \tilde{A}^* \) be a word. Then \( w \) labels a path starting at \( e \) if and only if \( ew \in R e \) in \( I \). The endpoint of such a path is \((ew)^{-1}(ew)\).

**Proof.** First suppose that \( w \) labels such a path; we proceed by induction on \( |w| \), the case \( |w| = 0 \) being clear. Suppose \( w = ux \) with \( x \in A \) and \( u \in A^* \).

Since \( u \) labels a path starting at \( e \), we have, by induction, that \( eu \in R e \) and the path ends at \( f = (eu)^{-1}eu \). It now follows that the last edge of the path labeled by \( w \) from \( e \) is \( (f,x) \) whence \( fx \in R f \) and the endpoint is \((fx)^{-1}fx \).

We can conclude that \( uf \in R uf \); but \( uf = uu^{-1}eu = eu \) whence we have

\[
ew = eux \in R eu \in R e.
\]
Also,
\[(fx)^{-1}fx = (u^{-1}eux)^{-1}(u^{-1}eux)\]
\[= x^{-1}u^{-1}eux^{-1}eux = w^{-1}ew = (ew)^{-1}(ew)\]
and the result follows.

Suppose now that \(w\) is such that \(ew \mathcal{R} e\). We induct on \(|w|\), the case \(|w| = 0\) being clear. Suppose \(w = ux\) with \(x \in \tilde{A}\) and \(u \in \tilde{A}^*\). Then \(e \mathcal{R} ew \leq \mathcal{R} eu \leq \mathcal{R} e\) whence \(eu \mathcal{R} e\) and so, by induction, \(u\) labels a path \(p\) from \(e\) with endpoint \((eu)^{-1}eu = u^{-1}eu\). But
\[u^{-1}eux = u^{-1}ew \mathcal{R} u^{-1}e \mathcal{R} u^{-1}eu\]
so we can add an edge \(((eu)^{-1}eu, x)\) to \(p\) to get a path from \(e\) labeled by \(w\).

One can verify, as above, that the endpoint of \(((eu)^{-1}eu, x)\) is \(((ew)^{-1}ew)\).

\[\square\]

**Corollary 8.2.** Let \(R' = \{uv^{-1} \mid u = v \in R\}\) (note that we do not reduce the word \(uv^{-1}\)); then every path in \(X\) labeled by an element of \(R'\) is a closed path.

**Proof.** Suppose \(w \in R'\) labels a path from a vertex \(e\). Then, by Proposition 8.1, \(ew \mathcal{R} e\) and the path ends at \((ew)^{-1}ew\). But \(w \in \mathcal{E}(I)\), whence \(ew \in \mathcal{E}(I)\). Since \(\mathcal{R}\)-equivalent idempotents are equal, it follows \((ew)^{-1}ew = ew = e\).

We now attach a 2-cell for every path labeled by an element of \(R'\) (such paths are closed by Corollary 8.2). Clearly the restriction of a defining path is a defining path and so \(X\) is an ordered 2-complex. Alternatively, one can describe the 2-cells as follows: For each relation \(u = v \in R\), whenever we find a vertex \(w\) from which both \(u\) and \(v\) label a path (in which case both these paths have the same terminus), we add a cell with boundary \(uv^{-1}\).

If \(I\) is a group, then \(X\) is the standard 2-complex of the corresponding group presentation \(\langle X \mid R' \rangle\) (again not reducing the elements of \(R'\)); here, there may be some spheres with boundaries \(xx^{-1}\) or \(x^{-1}x\), where \(x\) is a generator, as we use the same relations for the inverse semigroup and group presentations).

Viewing \(I\) as an inductive groupoid, we can define an ordered graph morphism \(\psi : X \to I\) by the identity on vertices and on edges by \(\psi(e, x) = ex\).

**Proposition 8.3.** The map \(\psi\) is an ordered graph morphism. Furthermore, any defining path is sent to an idempotent (that is, a local identity).

**Proof.** First we show that \(d\) is preserved. Indeed, if \((e, x) \in E(X)\), then \(ex \mathcal{R} e\) whence \(d(ex) = e = d(e, x)\). As to inverses,
\[\psi((ex)^{-1}ex, x^{-1}) = x^{-1}ex^{-1} = x^{-1}xx^{-1} = x^{-1}e = (ex)^{-1}.
\]
Order is preserved since the natural partial order on an inverse semigroup is compatible with multiplication. An easy induction argument shows that
It now follows that \( \psi \) induces a map \( \overline{\psi} : \Pi_1(X) \to I \). We wish to show that this is an isomorphism. Define a map \( \tau : A \to \Pi_1(X) \) (where we now view the latter as an inverse semigroup) by \( x \mapsto (xx^{-1}, x) \). If \( \tau \) induces a morphism from \( I \), then, since \( \overline{\psi} \tau \) is the identity on \( A \), it would have to induce an isomorphism.

**Lemma 8.4.** The map induced by \( \tau \) from \( \tilde{A}^+ \) to \( \Pi_1(X) \) takes a word \( w \) to the path labeled by \( w \) from \( uw^{-1} \) to \( w^{-1}w \).

**Proof.** First we observe that, since \( uw^{-1}w R w \), Proposition 8.1 shows that \( w \) labels a path from \( uw^{-1} \) to \( w^{-1}w \). We prove the result by induction. The case \( |w| = 1 \) follows from the definition. If \( w = ux \) with \( u \in \tilde{A}^+ \) and \( x \in \tilde{A} \), then, by induction, \( u \) maps to the path \( p \) labeled by \( u \) from \( uu^{-1} \) to \( u^{-1}u \). If \( u^{-1}u R u^{-1}ux \), then

\[
\begin{align*}
\tau & = uu^{-1}ux = w,
\end{align*}
\]

so there is an edge \((u^{-1}u, x)\), and \( w \) maps to \( p(u^{-1}u, x) \) which is a path from \( uu^{-1} = ww^{-1} \) to \((u^{-1}ux)^{-1}u^{-1}ux = w^{-1}w \) labeled by \( w \). If \( u^{-1}ux \not< R u^{-1}u \), then the image of \( w \) is the product of the path labeled by \( u \) ending at \( u^{-1}uxx^{-1} \) with the edge \((u^{-1}uxx^{-1}, x)\). But Proposition 8.1 shows that the path labeled by \( u^{-1} \) from \( u^{-1}uxx^{-1} \) ends at

\[
(xx^{-1}u^{-1})^{-1}xx^{-1}u^{-1} = uxx^{-1}u^{-1} = w^{-1}.
\]

The result follows. \( \square \)

**Proposition 8.5.** The map \( \tau \) induces a morphism from \( I \) to \( \Pi_1(X) \).

**Proof.** Suppose \( u = v \in R \). Let \( f = uu^{-1} = vv^{-1} \); then \( f = wv^{-1} = f^{-1}wv^{-1} \). It follows, by Proposition 8.1, that \( wv^{-1} \) labels a path from \( f \) to \( f \) in \( X \) which, by construction, is a defining path. Thus the paths from \( f \) labeled by \( u \) and \( v \) exist and are homotopic. But these paths are the images of the words \( u \) and \( v \) in \( \tilde{A}^+ \) by Lemma 8.4. It follows that \( \tau \) induces a morphism. \( \square \)

**Corollary 8.6.** Let \( I = \text{Inv} \langle A| R \rangle \) be an inverse semigroup presentation and \( X \) the standard ordered 2-complex for this presentation. Then \( I = \Pi_1(X) \).

Keeping our previous notation, the maximal group image of \( I \) is the group \( G \) presented by \( \langle A|R' \rangle \). Furthermore, there is a natural morphism \( \psi \) from
$X$ to the standard 2-complex $Y$ for $G$ with the above presentation (again, we do not reduce the words in $R'$ and we may have sphere representing the trivial relations: $xx^{-1} = 1 = x^{-1}x$). This map collapses all the vertices and projects to the second coordinate on edges. The 2-cells are mapped in the natural way. This map is, in fact, an immersion [20] on the 1-skeleton; also, any based 2-cell has at most one based lift to any vertex. McAlister’s $P$-theorem [10] can easily be shown to be equivalent to stating that $I$ is $E$-unitary (that is, the natural projection to $G$ is idempotent pure) if and only there is an ordered 2-complex $Z$ containing $X$ and an extension of $\psi$ to $Z$ which is a covering such that $\Pi_1(Z)$ is an enlargement of $\Pi_1(X)$ in the sense of [7, 21]. In fact, all the results of [21] can be restated and proved more generally in the context of ordered 2-complexes where the role of the derived ordered groupoid is replaced by what is called in homotopy theory, the mapping fiber. More on this will appear in a future paper.

We now turn to some examples.

**Example 8.7.** Let $I$ be the free inverse semigroup given by the presentation $\text{Inv} \langle A \rangle$. Then the standard ordered 2-complex $X$ consists of all finite subtrees of the Cayley graph of a free group on $A$ (with respect to generators $A$) containing 1 as a vertex (called *Munn trees*). We will denote a vertex by a pair $(T, g)$ where $T$ is a Munn tree and $g \in G$ labels a vertex of $T$. Edges will be denoted by triples $(T, g, x)$ where $T$ is a Munn tree, $g \in G$ is a vertex of $T$, and $x \in A$ labels an edge in $T$ from $g$. The domain of $(T, g, x)$ is $(T, g), (T, g, x)^{-1} = (T, gx, x^{-1})$. The order on $V(X)$ is given by $(T, g) \leq (T', g')$ if and only if $g = g'$ and $T' \subseteq T$. The order on $E(X)$ is given by $(T, g, x) \leq (T', g', x')$ if and only if $g = g', x = x'$, and $T' \subseteq T$. Note that $X$ is $\pi_1$-trivial.

**Example 8.8.** A *Brandt* semigroup is an inverse semigroup whose underlying set is $(J \times G \times J) \cup 0$, with $G$ a group and $J$ a set, and whose nonzero multiplications are given by

$$(i, g, j)(j, g', i') = (i, gg', ii').$$

This inverse semigroup will be denoted $\mathcal{M}^0(J, G)$. Suppose $J$ has a distinguished element $j_1$. If $G = \langle A | R \rangle$ is a presentation for $G$, we can obtain a presentation $\mathcal{M}^0(J, G)$ by taking as generators $J \setminus j_1 \cup A \cup z$ where, for $j \in J \setminus j_1$, $j \mapsto (j, 1, j_1)$, for $a \in A$, $a \mapsto (j_1, a, j_1)$, and $z$ maps to 0. Choose a distinguished element $a_0 \in A$. For relations, we use: The relations of $R$ but where 1 is replaced by $a_0a_0^{-1}$; the relations $j^{-1}j = a_0a_0^{-1}, j \in J \setminus j_1$; $j^{-1}j' = z$ if $j \neq j' \in J \setminus j_1, xz = z, x \in A \cup J \setminus j_1, a_0a^{-1} = a^{-1}a, a \in A$; $aa^{-1} = a_0a_0^{-1}, a \in A$; and the relations needed to make $z$ a zero. The standard ordered 2-complex $X$ has two connected components. One component has a single vertex 0 and edges which are loops labeled by $A \cup J \setminus j_1 \cup z$. The 2-cells give us the standard 2-complex for $G$ plus some 2-cells with
contractible defining paths and some 2-cells with defining paths homotopic to the various edges and their conjugates by $z$ (whence the local group is trivial). The other component has vertices of the form $(j, 1)$ with $j \in J$. At the vertex $(j_1, 1, j_1)$ there is a copy of the standard 2-complex of $G$ plus some 2-cells with contractible defining paths. For $j \in J \setminus j_1$, there is an edge from $(j, 1, j)$ to $(j_1, 1, j_1)$ labeled by $j$. There are also 2-cells with contractible defining paths labeled by $j^{-1}ja_0a_0^{-1}$. This connected component is clearly homotopy-equivalent to a wedge of the standard 2-complex for $G$ and some spheres, so the local group is $G$.

**Example 8.9.** Consider the bicyclic semigroup $I = \text{Inv} \langle x | xxx^{-1} = x \rangle$. One can check that standard ordered 2-complex has 1-skeleton: The interval $[0, \infty)$ with its usual graph structure; the vertices are ordered by $\geq$; the edges are incomparable. Each interval of length two is the seam glued of an attached American football. Thus the complex is homotopy-equivalent to a countably infinite wedge of spheres, and hence has trivial fundamental group.

Let $X$ be the standard ordered 2-complex of an inverse semigroup presentation $I = \text{Inv} \langle A | R \rangle$. We now consider the 1-skeleton of the Schützenberger complex $\tilde{X}$. The vertices are the elements of $I$. Edges are of the form $(s, (e, x))$ where $s^{-1}s = e$ and $(e, x)$ is an edge of $X$. Thus edges can be more succinctly described as pairs $(s, x)$ with $s \in I$ and $s^{-1}sx \mathcal{R} s^{-1}s$ or, equivalently, $sx \mathcal{R} s$; $d(s, x) = s$, $(s, x)^{-1} = (sx, x^{-1})$. Connected components correspond to $\mathcal{R}$-classes and 2-cells are added whenever an element of $R'$ labels a path. The reader familiar with the work of Stephen [24] will see immediately that the 1-skeleton of the connected component corresponding to an $\mathcal{R}$-class is the Schützenberger graph of that $\mathcal{R}$-class. Stephen’s Todd-Coxeter-like procedure to construct the Schützenberger graph can easily be modified to construct the corresponding component of the Schützenberger complex by adding a 2-cell every time an expansion is performed. We note that the left action of $I$ on $\tilde{X}$ is “essentially” the left Preston-Wagner representation [7].

The covering $\varphi : \tilde{X} \to X$ takes a vertex $s$ to $s^{-1}s$ and takes the connected component of $\tilde{X}_e$ corresponding to an idempotent $e$ onto the $D$-class $D_e$ of $X$ corresponding to the $D$-class of $e$ in $I$. Two elements of the $\mathcal{R}$-class of $e$ get identified if and only if they are $\mathcal{H}$-equivalent. In fact, $H_e$, the $\mathcal{H}$-class of $e$, acts freely on the left of $\tilde{X}_e$ and the quotient is $D_e$ by Proposition 7.3. Thus we have the following result:

**Theorem 8.10.** Suppose that $I$ is a finitely generated (presented) inverse semigroup and $H_e$ is a maximal subgroup with identity $e$ such that the left action of $H_e$ on its $\mathcal{R}$-class $R_e$ has finitely many orbits. Then $H_e$ is finitely generated (presented).
Proof. Let $X$ be the standard ordered 2-complex and $D_e \subseteq X$ the $D$-class of $e$. Our above discussion shows that $V(D_e) = H_e/R_e$ and hence is finite. Since $I$ is finitely generated, it follows that $D_e^{(1)}$ is a finite vertex, locally finite graph: That is, $D^{(1)}$ is finite. Thus $H_e = \pi_1(D_e)$ is finitely generated. If $I$ is also finitely presented, then $R'$ (defined as above) is finite. Since $D_e^{(1)}$ is also finite, we can only find finitely many paths in $D_e$ labeled by elements of $R'$ whence, by construction of $X$, $D_e$ is a finite 2-complex. It follows $H_e = \pi_1(D_e)$ is finitely presented. \qed

It is a simple exercise to give an explicit presentation of $H_e$ using the standard techniques for computing the fundamental group of a finite 2-complex. Similar results were obtained in [17] for semigroup presentations of regular semigroups using Reidemeister-Schreier-type rewriting systems.

This result leads us to define a finitely generated inverse semigroup to be hyperbolic if $\tilde{X}$ is a Gromov hyperbolic (disconnected) space, and all the maximal subgroups act co-compactly on the corresponding component of $\tilde{X}$. This would be a generalization of hyperbolic groups and an immediate consequence of the definition is that all the maximal subgroups are hyperbolic (since any group acting properly discontinuously and co-compactly on a hyperbolic space is hyperbolic). Similarly, we define a finitely presented inverse semigroup to be Fuchsian if its Schützenberger complex is planar (after identifying 2-cells which give rise to the same defining paths, and removing 2-cells with contractible defining paths), and each maximal subgroup acts co-compactly on the corresponding component of $\tilde{X}$. Once again, this implies that the maximal subgroups are Fuchsian. These classes will be studied in future papers; also notions of Van Kampen diagrams over inverse semigroup presentations will be introduced in future work.

We end this section with the following observation: If $D$ is a $D$-class of $I$, then $I$ acts on the right of the set of $\mathcal{L}$-classes of $D$ by partial bijections. The transition inverse semigroup is what Rhodes calls the right letter mapping semigroup of $I$ corresponding to $D$ [6]. One can then form the right letter mapping graph of $D$ whose vertices are the $\mathcal{L}$-classes of $D$ and whose edges are of the form $(b, x)$ whenever $bx$ is defined in the right letter mapping transformation semigroup. One has $d(b, x) = b$ and $(b, x)^{-1} = (bx, x^{-1})$. It is straightforward to verify that the map taking $b$ to the unique idempotent in $b$ is an isomorphism of the right letter mapping graph with the $D$-class of $X$ corresponding to $D$. The hypothesis of Theorem 8.10 can be rephrased in terms of asking that the right letter mapping transformation semigroup of the $D$-class of $e$ be finite. Since the Munn representation is the direct sum of the right letter mapping representations, $X^{(1)}$ can be viewed as the graph of the Munn representation. Similarly, $\tilde{X}^{(1)}$ can be view as the graph of the right Preston-Wagner representation which is the direct sum of the Schützenberger representations.

In this section, we study amalgams of the various structures we have been considering.

9.1. Amalgams of partially ordered sets. An amalgam of partially ordered sets is a triple of partially ordered sets \((S,U,T)\) with \(S \cap T = U\). The amalgamated product \(S \coprod_U T\) is the partially ordered set \(S \cup T\) with order the union of the orders of \(S\) and \(T\) and, in addition, for \(s \in S\) and \(t \in T\), one defines \(s \leq t\) if there exists \(u \in U\) with \(s \leq u \leq t\) (note that if \(s\), respectively, \(t\) is in \(U\), this situation occurs if and only if \(s \leq t\) in \(S\), respectively, \(T\)); \(t \leq s\) is defined similarly.

**Proposition 9.1.** \(S \coprod_U T\) is a partially ordered set and, given any partially ordered set \(P\) and morphisms \(\varphi : S \to P\), \(\psi : T \to P\), agreeing on \(U\), there is a unique morphism \(\tau : S \coprod_U T \to P\) extending \(\varphi\) and \(\psi\).

*Proof.* For anti-symmetry, it suffices to show that, for \(s \in S\) and \(t \in T\), \(s \leq t\) and \(t \leq s\) implies \(s = t\). Indeed, we then have \(u, u' \in U\) with \(s \leq u \leq t\) and \(t \leq u' \leq s\); so \(s \leq u \leq u' \leq s\) and \(u \leq t \leq u'\) whence \(s = u = u' = t\). As for transitivity, the only cases to deal with (up to exchanging the role of \(s\) and \(t\)) are: \(s \leq s', s' \leq t; s \leq t, t \leq t';\) and \(s \leq t, t \leq s'\) where \(s, s' \in S\) and \(t, t' \in T\). In the first case, we have \(s \leq s' \leq u \leq t\) for some \(u \in U\) whence \(s \leq t\); the second case is similar. For the third case, we have \(u, u' \in U\) such that \(s \leq u \leq t\) and \(t \leq u' \leq s'\) whence \(u \leq u'\) and \(s \leq u \leq u' \leq s'\).

Now set \(\tau = \varphi \cup \psi\). To see that \(\tau\) preserves order, it is enough, without loss of generality, to show that, for \(s \in S\), \(t \in T\), if \(s \leq t\) then \(\tau(s) \leq \tau(t)\). But in this case, we have \(u, u' \in U\) with \(s \leq u \leq t\) whence

\[
\tau(s) = \varphi(s) \leq \varphi(u) = \psi(u) \leq \psi(t) = \tau(t)
\]

and the result follows. \(\square\)

**Proposition 9.2.** Suppose \((S,U,T)\) is a partially ordered set amalgam with \(U\) an order ideal of \(S\) and \(T\). Then \(S \coprod_U T\) is just \(S \cup T\) with the union of the two partial orders, and \(S\), \(T\), and \(U\) are order ideals.

*Proof.* We first prove the statement about order ideals. Suppose \(x \leq s \in S\) and \(x \in T\). Then there exists \(u \in U\) with \(x \leq u \leq s\) and so \(x \in U \subseteq S\). Thus \(S\) and, by a dual argument, \(T\) are order ideals. Since \(U = S \cap T\), \(U\) is also an order ideal. It now follows immediately that the order is just the union of the orders. \(\square\)

We now give a condition under which the partially ordered set amalgamation of a semilattice amalgam is a semilattice.

**Proposition 9.3.** Suppose \((S,U,T)\) is a partially ordered set amalgam with \(U\) an order ideal of \(S\) and \(T\). Suppose, further, that \(S\) and \(T\) are semilattices and that \(U\) has a maximum \(e\). Then \(S \coprod_U T = S \cup T\) is a semilattice.
Proof. Note first that, since $U$ is an order ideal in $S$ and $T$, it is a semilattice. By Proposition 9.2, it suffices to show that if $s \in S$ and $t \in T$, then $s$ and $t$ have a meet. We claim that the meet is $(s \wedge e) \wedge (e \wedge t)$; this element is well-defined since $U$ is an order ideal in $S$ and $T$. Indeed, any common lower bound $g$ is in $S \cap T = U$. Thus $g \leq s, t, e$ whence $g \leq s \wedge e, e \wedge t \in U$ and the result follows.

$\square$

9.2. Amalgams of ordered 2-complexes. An amalgam of ordered 2-complexes consists of a triple of ordered 2-complexes $(X, Y, Z)$ with $Y = X \cap Z$. The amalgamated product $X \coprod_{Y} Z$ is defined to be the universal ordered 2-complex with morphisms of $X$ and $Z$ into it agreeing on $Y$ (if such exists).

Suppose $(X, Y, Z)$ is an amalgam of ordered 2-complexes. Then $X \cup Z$ has the natural structure of a 2-complex with ordered vertex and edge sets; that is:

1. $V(X \cup Z) = V(X) \coprod_{V(Y)} V(Z)$;
2. $E(X \cup Z) = E(X) \coprod_{E(Y)} E(Z)$;
3. $C(X \cup Z) = C(X) \cup C(Y)$.

There are naturally induced, order-preserving maps $d$ and $e \mapsto e^{-1}$ by the universal property of the amalgamated product of partially ordered sets (also $e \neq e^{-1}$ if such is true for $X$ and $Z$). We call this the natural order structure on $X \cup Z$.

We say an amalgam of ordered 2-complexes is tame if $X \cup Z$ with the natural order structure is an ordered 2-complex. It is immediate that in this case $X \coprod_{Y} Z = X \cup Z$. We now give some examples of tame amalgams.

Proposition 9.4. Suppose $(X, Y, Z)$ is an ordered graph amalgam such that $V(Y)$ is an order ideal of $V(X)$ and $V(Z)$. Then $(X, Y, Z)$ is tame.

Proof. We first show that $(X \cup Y)^{(1)}$ is an ordered graph. Suppose $v \leq w$ and $e \in \text{Star}(w)$; we must show there is a unique edge $e' \leq e$ in $\text{Star}(v)$. Without loss of generality, assume $e \in E(X)$ (and so $w \in V(X)$). Then, since $V(X)$ is an order ideal by Proposition 9.2, $v \in V(X)$ and $(v|e)$ is one such edge. Suppose $e'$ is another such (necessarily $v \in V(Y)$ and $e' \in E(Z)$). But then there is an edge $e'' \in E(Y)$ with $e' \leq e'' \leq e$, so $e'' = (v|e'') \in E(Y) \subseteq E(X)$ and $e' \leq e$ whence $e' = (v|e)$ as desired. Note that the above argument shows that $E(X)$ and $E(Z)$ are order ideals in $E(X \cup Z)$.

Assume now that $p$ is a defining path of $(c, f_{c}) \in C(X) \cup C(Y)$ and $v \leq d(p)$. Without loss of generality, assume $(c, f_{c}) \in C(X)$. Then $v \in V(X)$ (since $V(X)$ is an order ideal), and the restriction $(v|p)$ in $X$ and in $X \cup Z$ coincide whence $(v|p)$ is a defining path of a 2-cell in $X$.

In particular, all amalgams of unordered groupoids are tame.

An ordered subcomplex (subgroupoid) $Y$ of an ordered 2-complex (groupoid) $X$ is said to be $D$-saturated in $X$ if it is a union of $D$-classes of $X$. 
Proposition 9.5. Let \((X, Y, Z)\) be an amalgam of inductive groupoids such that \(Y\) is \(D\)-saturated in both \(X\) and \(Z\). Then \((X, Y, X)\) is tame.

Proof. We begin by showing that \((X \cup Y)\) is an ordered graph. Suppose, without loss of generality, that \(e \in E(X)\) and \(v \leq d(e)\). If \(v \in V(X)\), then \((v|e) \leq e\) and if \(f \leq e\) with \(d(f) = v\) and \(f \neq (v|e)\), then \(f \in E(Z)\) and \(v \in V(Y)\). But \(Y\) is \(D\)-saturated in \(Z\), so \(f \in E(Y)\) whence \(f = (v|e)\) proving uniqueness in this case. Suppose now \(v \in V(Z) \setminus V(Y)\), then there exists \(w \in V(Y)\) with \(v \leq w \leq d(e)\). Then, since \(Y\) is \(D\)-saturated in \(X\), \((w|e) \in Y\) whence \(f = (v|(w|e)) \leq e\) and \(d(f) = v\). Suppose now that \(f' \leq e\) with \(d(f') = v\). Then \(f' \in E(Z)\) and there exists \(y \in E(Y)\) with \(f' \leq y \leq e\). Let \(u = w \wedge d(y) \in V(Y)\). Then \((u|y) = (u|e) = (u|(w|e))\) whence
\[f' = (v|y) = (v|(u|e)) = (v|(w|e))\]
and we have uniqueness.

Now suppose \(p\) is the defining path of \((c, f_c) \in C(X) \cup C(Z)\); without loss of generality, we take \((c, f_c) \in C(X)\). Suppose \(v \leq d(p)\). If \(v \in V(X)\), then \((v|p)\) is the same in \(X\) and \(X \cup Z\), so it is a defining path of some 2-cell in \(C(X)\). Suppose now \(v \in V(Z) \setminus V(Y)\). Then there exists \(w \in V(Y)\) with \(v \leq w \leq d(p)\). Since \(Y\) is \(D\)-saturated in \(X\), \((w|p)\) is in \(Y\) and is the defining path of a 2-cell of \(X\) which must, in fact, belong to \(Y\) since \(Y\) is \(D\)-saturated. Then \((v|(w|p))\) is the defining path of a 2-cell in \(Z\). \(\square\)

9.3. Graphs of groups. To study the maximal subgroups of an amalgamated product of ordered groupoids, we need to recall the notion of a graph of groups [19]. A graph of groups \((G, X)\) consists of a graph \(X\) and an assignment to each vertex \(v\) of a group \(G_v\), to each pair of edges \(\{e, e^{-1}\}\) of a group \(G_e = G_{e^{-1}}\), and to each edge \(e\) an inclusion \(\iota_e : G_e \to G_{d(e)}\). If \((G, X)\) is a connected graph of groups and \(T\) is a maximal subtree of \(X\), then fundamental group of \((G, X)\) with respect to the tree \(T\) is
\[\pi_1(G, X) = \langle G_v \ (v \in V(X)), y_e \ (e \in E(X)) | y_t = 1 \ (t \in T), y_e^{-1} y_{e^{-1}} y_e = \iota_{e^{-1}}(g) y_e = \iota_e(g) \ (e \in E(X), g \in G_e) \rangle.\]

This group is independent of the choice of \(T\) [19].

A graph of 2-complexes \((C, X)\) consists of a graph \(X\) and an assignment to each vertex \(v\) of a 2-complex \(X_v\) (called a vertex complex), to each pair of edges \(\{e, e^{-1}\}\) of 2-complex (called an edge complex) \(X_e = X_{e^{-1}}\), and to each edge \(e\) an inclusion \(\iota_e : X_e \to X_{d(e)}\). The realization of \((C, X)\) is obtained by replacing each vertex \(v\) by \(X_v\) and each pair of edges \(\{e, e^{-1}\}\) by \(X_e \times [0, 1]\); one then glues, for each pair of edges \(\{e, e^{-1}\}\), \(X_e \times 0\) to the copy of \(X_{e^{-1}}\) in \(X_{d(e)}\) and \(X_e \times 1\) to the copy in \(X_{t(e)}\). In general, we say that a graph of 2-complexes \(X\) gives a graph of 2-complexes decomposition of a 2-complex \(Y\) if \(Y\) is homotopy-equivalent to the realization of \((C, X)\).
A simple application of the Seifert-Van Kampen Theorem shows that if \( X \) is the geometric realization of a connected graph of connected 2-complexes, then \( \pi_1(X) \) is isomorphic to the fundamental group of the graph of groups obtained by replacing each vertex and edge complex by its fundamental group. Conversely, the fundamental group of any connected graph of groups can be so realized.

The following simple observation, which we state as a proposition, will be key to studying the structure of maximal subgroups of amalgams:

**Proposition 9.6.** Let \((C, X)\) be a graph of 2-complexes. Then its realization has the following decomposition as a graph \((C', Y)\) of connected 2-complexes: \( V(Y) \) is the set of \( D \)-classes of the vertex complexes of \( X \); \( E(Y) \) is the set of \( D \)-classes of the edge complexes of \( X \); if \( e \in E(X) \) and \( Z \) is a \( D \)-class of \( X_e \), then \( Z \) is a subcomplex of a unique \( D \)-class of \( X_{d(e)} \) which we define to be the domain of the edge corresponding to \( Z \); the inverse of the edge corresponding to \( Z \) is the copy of \( Z \) in \( X_e^{-1} \); and the vertex (edge) complex corresponding to a \( D \)-class is the \( D \)-class itself.

We illustrate the above proposition in Figure 1 where we have a segment of disconnected complexes as this is the case of interest.

![Figure 1](image_url)

**Figure 1.** An illustration of Proposition 9.6.

Here, the left vertex complex (surrounded in large thick dashes) has two \( D \)-classes \( x1 \) and \( x2 \) (dashed lines); the right vertex complex (surrounded in small thick dashes) has three \( D \)-classes \( y1 \), \( y2 \), and \( y3 \) (dashed lines); the edge complex (surrounded in thick solid lines) has four \( D \)-classes \( z1 \),
z2, z3, and z4 (dotted lines). The resulting decomposition is a connected, bipartite graph of connected 2-complexes with five vertices and four edges (see Figure 2); this will always be the case for a segment of 2-complexes although in general the bipartite graph will not be connected.

![Figure 2. The underlying graph in Figure 1.](image)

Proposition 9.6 gives us a means to obtain a graph of groups decomposition of any maximal subgroup of the fundamental groupoid of a 2-complex with a graph of 2-complexes decomposition. The main such decomposition of interest in the is paper is the following:

**Proposition 9.7.** Let \((X, Y, Z)\) be a tame amalgam of ordered 2-complexes. Then \(X \coprod_Y Z\) is homotopy-equivalent to the realization of the graph of 2-complexes \((\mathcal{C}, \mathcal{W})\) whose underlying graph is a segment with \(X\) as the left vertex complex, \(Z\) as the right vertex complex, and \(Y\) as the edge complex.

**Proof.** By assumption \(X \coprod_Y Z = X \cup Z\). The homotopy equivalence is induced by applying to the realization of \((\mathcal{C}, \mathcal{W})\) the contraction of \(Y \times [0, 1]\) to \(Y \times 0\). □

**9.4. Amalgams of ordered groupoids.** An *amalgam* of ordered groupoids consists of a triple \((S, U, T)\) such that \(U = S \cap T\). The amalgamated product \(S \coprod_U T\) is then the universal object with maps of \(S\) and \(T\), agreeing on \(U\), into it (if such exists). Our goal is to study topologically the structure of the maximal subgroups of the amalgamated product. Our results generalize those of [3, 12] for full amalgams of regular semigroups, but our techniques are different.

If \((S, U, T)\) is an ordered groupoid amalgam, we define a graph of groups \((\mathcal{G}_{(S, U, T)}, \Gamma)\) as follows: \(V(\Gamma)\) is the union of the set of \(D\)-classes of \(S\) and the set of \(D\)-classes of \(T\); \(E(\Gamma)\) is the set of \(D\)-classes of \(U\) (and formal inverses of such); each \(D\)-class of \(U\) is contained in a unique \(D\)-class of \(S\), which we take as the initial vertex of the corresponding edge, and a unique \(D\)-class of \(T\), which we take as its final vertex; we associate to each vertex and edge the maximal subgroup of the corresponding \(D\)-class with the obvious inclusion maps (cf. [3]). If \(v \in V(S) \cup V(T)\), we let \((\mathcal{G}_v, \Gamma_v)\) be the connected component of \((\mathcal{G}_{(S, U, T)}, \Gamma)\) containing the \(D\)-class of \(v\).
We say an ordered 2-complex amalgam \((X_1, X_2, X_3)\) represents an ordered groupoid amalgam \((S_1, S_2, S_3)\) if \(\Pi_1(X_i) = S_i, \ i = 1, 2, 3.\) Abstract nonsense shows that if \(X_1 \coprod X_2 \coprod X_3\) exists, then \(S_1 \coprod S_2 \coprod S_1 = \Pi_1(X_1 \coprod X_2 \coprod X_3).\) We say that an ordered groupoid amalgam is tame if it can be represented by a tame ordered 2-complex amalgam.

**Theorem 9.8.** Suppose \((S, U, T)\) is a tame ordered groupoid amalgam. Then \(S \coprod U \coprod T\) exists, has vertex set \(V(S) \cup V(T),\) and the maximal subgroup corresponding to a vertex \(v\) is \(\pi_1(G_v, \Gamma_v).\)

**Proof.** Choose a tame ordered 2-complex amalgam \((X, Y, Z)\) representing \((S, U, T).\) Then, \(X \coprod Y \coprod Z = X \cup Z\) has a graph of 2-complexes decomposition as per Proposition 9.7 whence Proposition 9.6 shows that the maximal subgroups are as described in the theorem statement. \(\square\)

Note that the theorem always applies if the groupoids are unordered.

**Corollary 9.9.** Suppose \((S, U, T)\) is a tame ordered groupoid amalgam such that \(U\) has only trivial subgroups. Then the local groups of \(S \coprod U \coprod T\) are free products of local groups of \(S,\) local groups of \(T,\) and free groups. In particular, if all the local groups of \(S\) and \(T\) are trivial, then the subgroups of \(S \coprod U \coprod T\) are free.

**Proof.** The fundamental group of a connected graph of groups in which all edge groups are trivial is a free product of vertex groups and a free group. \(\square\)

**Proposition 9.10.** Let \((S, U, T)\) be an ordered groupoid amalgam. Suppose that either \(V(U)\) is an order ideal of \(V(S)\) and \(V(T),\) or \(U\) is \(\mathcal{D}\)-saturated in \(S\) and \(T,\) and \(S, U,\) and \(T\) are inductive. Then \((S, U, T)\) is tame.

**Proof.** Let \(X, Y, Z\) be the multiplication table presentations of, respectively, \(S, U, T.\) Then \((X, Y, Z)\) represents \((S, U, T).\) In the case that \(V(U)\) is an order ideal, we see that \((X, Y, Z)\) is tame by Proposition 9.4, while in the case that \(V(U)\) is \(\mathcal{D}\)-saturated and the amalgam consists of inductive groupoids, the result follows from Proposition 9.5. \(\square\)

Note that if \(S, U,\) and \(T\) are inductive and \(U\) is \(\mathcal{D}\)-saturated, then the graph of groups \((\mathcal{G}(S, U, T), \Gamma)\) consists of isolated vertices corresponding to \(\mathcal{D}\)-classes not in \(U\) and disjoint segments corresponding to \(\mathcal{D}\)-classes in \(U;\) the inclusions are isomorphisms for these segments. It follows that if \(v \in V(S) \cup V(T) = V(S \coprod U \coprod T),\) then the maximal subgroup at \(v\) is unchanged on forming the amalgamated product.

### 9.5. Amalgams of inverse and regular semigroups

We now wish to apply the above theory to inverse and regular semigroup amalgams. If \((S, U, T)\) is an amalgam of inverse (regular) semigroups, we use \(S \ast_U T\) for the amalgamated product. In the case of inverse semigroup amalgams \((S, U, T),\) a well-known theorem of Hall [4] shows that \(S\) and \(T\) embed in
$S \ast_U T$ in such a manner that their intersection is $U$. He later generalized this to full regular semigroup amalgams [5] (that is, amalgams $(S, U, T)$ such that $\mathcal{E}(S) = \mathcal{E}(U) = \mathcal{E}(T)$).

The following generalizes results of [3, 12] where the case of full amalgams is considered; see [3] for examples; other generalizations (some of which we shall obtain below) can be found in [1].

**Theorem 9.11.** Suppose $(S, U, T)$ is an inverse (regular) semigroup amalgam with $\mathcal{E}(S) = \mathcal{E}(U)$ an order ideal of $\mathcal{E}(T)$. Then $S \ast_U T = S \coprod_U T$, and the maximal subgroup corresponding to an idempotent $e \in \mathcal{E}(S \ast_U T) = \mathcal{E}(S) \cup \mathcal{E}(T)$ is $\pi_1(\mathcal{G}, \Gamma_e)$. In particular, if $S$ and $T$ only have trivial subgroups, then the subgroups of $S \ast_U T$ are free.

**Proof.** Since $\mathcal{E}(U)$ is an order ideal in $\mathcal{E}(S)$ and $\mathcal{E}(T)$, Proposition 9.10 shows $(S, U, T)$ is tame whence Theorem 9.8 applies. Since

$$V \left( S \coprod_U T \right) = V(S) \cup V(T) = V(T)$$

is a semilattice (biordered set), $S \coprod_U T$ is an $(r)$-inductive groupoid. Suppose $\varphi : S \to I$ and $\psi : T \to I$ are homomorphisms agreeing on $U$ where $I$ is an inverse (regular) semigroup. Then the induced ordered groupoid map $\tau : S \coprod_U T \to I$ agrees with $\psi$ on $V(T) = V(S \coprod_U T)$ and therefore preserves the semilattice (biordered) structure and so $\tau$ is a homomorphism of inverse (regular) semigroups.

The above theorem applies, in particular, to the case of full amalgams. Assume $(S, U, T)$ is a regular semigroup amalgam satisfying the hypotheses of Theorem 9.11. If $U$ has only trivial subgroups, then Corollary 9.9 and the Kurosh Theorem [8] imply that the subgroups of $S \ast_U T$ are isomorphic to free products of free groups and subgroups of the factors $S$ and $T$. Work of Ordman [13, 14] gives normal forms for amalgamated products of unordered groupoids. Since the above theorem implies that the underlying groupoid of $S \ast_U T$ is the amalgamated product in the category of unordered groupoids, his results (combined with our description of the order) can be used to completely understand $S \ast_U T$. Also, standard results on groupoid amalgams [13, 14] show that $S$ and $T$ embed in $S \ast_U T = S \coprod_U T$ with intersection $U$. Since the inclusions are the identity on vertices, they preserve the $(r)$-inductive structure. Thus we have the following generalization of a theorem of Hall [5]:

**Theorem 9.12.** Let $(S, U, T)$ be a regular semigroup amalgam with $\mathcal{E}(U) = \mathcal{E}(S)$ an order ideal in $\mathcal{E}(T)$. Then $S$ and $T$ embed in $S \ast_U T$ with intersection $U$. 
An illustrative example is the case where \( S = \text{Inv} \langle x \mid x^3 = 0 \rangle \), \( T \) is a copy of the subsemigroup of \( S \) generated by \( x^2 \), made disjoint off the idempotents, and \( U = E(T) \). Note that \( U \) is an order ideal in \( E(S) \). Using the multiplication table presentation, \( S \) can be represented by a 2-complex with three components: The top component is a segment of length 1; the middle component is a segment of length 2; and the bottom component is a regular 1-cell. One can similarly represent \( T \), only one drops the top component. Each of the original semigroups is finite with trivial subgroups. Then \( S \ast_U T \) is represented by identifying the two bottom components and gluing the two middle components along the vertices. It follows that the maximal subgroup corresponding to the middle component is free of rank 2 and the other maximal subgroups are trivial. We invite the reader to construct \((G, \Gamma)\) and verify that the answer obtained in this manner coincides.

We are now in a position to prove some more difficult results on amalgams which coincide in part with results of [1]. Fix a tame inverse semigroup amalgam \((S, U, T)\). The inclusions of \( S \) and \( T \) induce a natural ordered groupoid morphism \( \varphi : S \coprod_U T \to S \ast_U T \). Our goal is to show that this map is an embedding of \( S \coprod_U T \) as a \( D \)-saturated ordered subgroupoid.

**Lemma 9.13.** Let \((S, U, T)\) be a tame amalgam of inductive groupoids. Then the inclusions of \( S \) and \( T \) into \( S \coprod_U T \) are embeddings. Furthermore, if \( e, f \in E(S) \), then \( ef \) is the meet of \( e \) and \( f \) in \( S \coprod_U T \) and dually for \( e, f \in E(T) \).

**Proof.** By Theorem 9.8, the inclusion of \( S \) into \( S \coprod_U T \) is the identity on vertices whence it gives an idempotent-separating inverse semigroup homomorphism onto its image (which is an inductive ordered subgroupoid of \( S \coprod_U T \)). But, by Hall’s Theorem [4], the inclusion of \( S \) into \( S \ast_U T \) is an embedding. Since this inclusion factors through \( \varphi \), it follows that \( S \) embeds into \( S \coprod_U T \). The result for \( T \) is dual.

As to the second statement suppose \( g \in E(T) \setminus E(U) \) and \( ef \leq g \leq e, f \). Then there exist \( u, u' \in E(U) \) with \( g \leq u \leq e \) and \( g \leq u' \leq f \) whence \( ef \leq g \leq uu' \leq ef \). It follows that \( g = uu' \), a contradiction. We conclude \( ef \) is the meet of \( e, f \) in \( S \coprod_U T \); a dual argument holds if \( e, f \in E(T) \). \( \square \)

By [7, Theorem 4.1.9], there is an inductive groupoid \( I \) so that \( S \coprod_U T \) sits inside \( I \) as an ordered subgroupoid and the inclusion preserves whatever meets exist in \( S \coprod_U T \). It follows from the above lemma that \( S \) and \( T \) live as inverse subsemigroups of \( I \) whence there is a morphism \( \tau : S \ast_U T \to I \). Note that \( \tau \varphi \) is the identity on \( S \) and \( T \) and hence, by the universal property, on \( S \coprod_U T \).

**Lemma 9.14.** Let \((S, U, T)\) be a tame amalgam of inductive groupoids. Then \( \varphi|_{V(S \coprod_U T)} \) is an order embedding.
Lemma 9.15. Let \( (S, U, T) \) be a tame amalgam of inductive groupoids. Suppose \( e \in V(S \coprod_U T) \), \( y \in S \ast_U T \) and \( e \leq d(y) \). Then \( ey \in \varphi(S \coprod_U T) \).

Proof. We induct on the length of a factorization of \( y \) as a product of elements of \( S \cup T \). If \( y \) has length 1, then \( y \in S \cup T \). Since \( \varphi \) is an order embedding, \( e \leq y \) in \( S \coprod_U T \) whence \( (e|y) \) exists in \( S \coprod_U T \). But then \( \varphi(e|y) \leq y \) and \( d(\varphi(e|y)) = e \) so \( ey = \varphi(e|y) \). Suppose, without loss of generality, \( y = sz \) with \( s \in S \) and where \( z \) has factorization of length \( n - 1 \) as a product of elements of \( S \cup T \). Then \( e \leq d(y) \leq d(s) \), so, by the above case, \( es = \varphi(w) \). Note that \( ey = (es)z \) and \( d(ey) = e = d(es) \). Thus
\[
\begin{align*}
    r(es) &= s^{-1}es = s^{-1}d(es)z = (s^{-1}es)zz^{-1}(s^{-1}es) \leq d(z).
\end{align*}
\]
Since \( r(es) = r(\varphi(w)) = r(w) \in V(S \coprod_U T) \), \( r(w)z = \varphi(v) \) for some \( v \in S \coprod_U T \) by induction, and, necessarily, \( d(v) = r(es) = r(w) \). But then
\[
    ey = esz = esr(es)z = (es)(r(w)z) = \varphi(w)\varphi(v) = \varphi(wv)
\]
and the result follows. \( \Box \)

Theorem 9.16. Let \( (S, U, T) \) be a tame amalgam of inductive groupoids. Then \( S \coprod_U T \) embeds as a \( D \)-saturated ordered subgroupoid of \( S \ast_U T \). Thus if \( e \in E(S) \cup E(T) \), then the maximal subgroup at \( e \) is \( \pi_1(\mathcal{G}_v, \Gamma_v) \). This applies, in particular, if \( E(U) \) is an order ideal of \( E(S) \) and \( E(T) \), or \( U \) is \( D \)-saturated in \( S \) and \( T \).

Proof. Let \( K \) be the \( D \)-saturated subgroupoid of \( S \ast_U T \) consisting of the \( D \)-classes of elements of \( S \cup T \). Notice this is an ordered subgroupoid of \( S \ast_U T \) and that \( \varphi(S \coprod_U T) \subseteq K \). We show that \( \varphi : S \coprod_U T \rightarrow K \) is onto; it will then follow that \( \tau \) takes \( K \) to \( S \coprod_U T \), and that \( \varphi\tau|_K \) is the identity thereby proving the theorem.

First observe that to prove \( \varphi \) maps onto \( K \), it suffices to show that it is a fibration; indeed, any edge or vertex of \( K \) lies on a path starting at an element of \( E(S) \cup E(T) \) and this path can be lifted under a fibration. Suppose then that \( e \in E(S) \cup E(T) \) and \( x \in K \) with \( d(x) = e \); then \( e \leq d(x) \), so Lemma 9.15 shows that \( x = ex \in \varphi(S \coprod_U T) \).

Once again, this theorem completely describes the structure of the ordered subgroupoid of \( S \ast_U T \) consisting of the \( D \)-classes of idempotents of \( S \) or \( T \).
In the case that $U$ is $D$-saturated in $S$ and $T$ and $e \in \mathcal{E}(U)$, then the maximal subgroup at $e$ in $S$ and in $T$ are the same. The following corollary is then immediate from the remarks after Proposition 9.5:

**Corollary 9.17.** Let $(S,U,T)$ be an amalgam of inverse semigroups such that $U$ is $D$-saturated in $S$ and $T$. Then, for $e \in \mathcal{E}(S)$ (respectively, $\mathcal{E}(T)$), the maximal subgroups at $e$ in $S *_U T$ is the same as in $S$ (respectively, $T$).

The above situation occurs, for instance, if $S$ and $T$ have a common ideal $U$. Another possible application is when $S$ and $T$ have a common $D$-class $U$ which is a group (for instance, the minimal ideal and the group of units are always $D$-classes which are subgroups).

We end with a discussion of some constructions which are similar in spirit to amalgams and to which our techniques apply.

A map of inverse semigroups $\varphi : S \to T$ is said to be a prehomomorphism if $\varphi(s^{-1}) = \varphi(s)^{-1}$ and $\varphi(st) \leq \varphi(s)\varphi(t)$. It is shown \cite{7} that prehomomorphisms correspond to general ordered groupoid morphisms of inductive groupoids (as opposed to those preserving the meet). If $(S,U,T)$ is an inverse semigroup amalgam, then we define the pre-amalgamated product of $S$ and $T$ over $U$, denoted $S *'_U T$, to be the pushout \cite{9} of $S$ and $T$ over $U$ in the category of inverse semigroups and prehomomorphisms. It follows by considering the map from the pre-amalgamated product to the amalgamated product that $S$ and $T$ map injectively into the pre-amalgamated product with intersection $U$.

As a simple example, if $S$ and $T$ are inverse semigroups with 0 and $U = 0$, then $S *'_U T$ is the 0-direct union of $S$ and $T$, that is, the semigroup $S \cup T$ where all products of elements of $S$ with elements of $T$ are 0.

**Theorem 9.18.** Suppose $(S,U,T)$ is an amalgam of inverse semigroups such that $\mathcal{E}(U)$ is an order ideal of $\mathcal{E}(S)$ and $\mathcal{E}(T)$, and $\mathcal{E}(S) \cup \mathcal{E}(T)$ is a semilattice. Then $S *'_U T = S \coprod_U T$, the ordered groupoid amalgamated product of $S$ and $T$ over $U$, and the maximal subgroups can be described as per Theorem 9.8. Furthermore, the natural maps of $S$ and $T$ into $S *'_U T$ are homomorphisms.

**Proof.** First observe that the hypotheses of Theorem 9.8 hold since $(S,U,T)$ is tame by Proposition 9.10. Our assumptions imply that $V(S \coprod_U T)$ is a semilattice so $S \coprod_U T$ is an inductive groupoid. The desired universal property is then automatic. The last statement is clear from the construction of $S \coprod_U T$. \hfill \Box

Note that in the context of the above theorem, Theorem 9.16 implies that $S *'_U T$ lives in $S *_U T$ as a $D$-saturated ordered subgroupoid; but the inclusion does not in general preserve meets (think about the inclusion of $S *'_0 T$ into $S *_0 T$ for inverse semigroups $S$ and $T$ with zeroes). Proposition 9.3 allows us to find examples of when the theorem applies. One example is the case
where \( S \cap T \) is a common ideal \( U \) which is a monoid with identity \( e \). The case of the 0-direct union is such. It is easy to show that in this case \( S *_U T \) has the following structure: \( S *_U T = S \cup T \) with the product in \( S \) and \( T \) as usual, and with the product \( st = (se)(et) \) for \( s \in S, t \in T \). In this case, of course, the structure of the maximal subgroups is transparent: They are just the maximal subgroups of \( S \) and \( T \).

A more general example is if \( S \) and \( T \) have ideals \( U_S, U_T \) (respectively) which are monoids with a common full inverse submonoid \( U \) (so \( E(U_S) = E(U) = E(U_T) \)). In this case, the theorem really does tell us something about the structure of the maximal subgroups. The simplest example is when \( S \) and \( T \) have minimal ideals and \( U \) is a common subgroup of the minimal ideals.

We mention that Yamamura [25] defines a notion of a graph of full inverse monoids and the fundamental inverse monoid of such. Full amalgams and full HNN-extensions are special cases of this construction. In the paper, he obtains a graph of groups decomposition of the maximal subgroups of such a fundamental inverse monoid similar to that for full amalgams. We leave it as an exercise for the reader to obtain these results by realizing the fundamental inverse monoid as the fundamental ordered groupoid of an ordered 2-complex with an obvious graph of 2-complexes decomposition and then using Theorem 9.6.

We conjecture that actions of inverse semigroups on ordered forests, as per [25], can be completely classified using ordered 2-complexes and the usual translation between Bass-Serre theory and graphs and trees of 2-complexes. In particular, we guess the solution will obtain the structure of such as the fundamental ordered groupoid of an ordered graph of groups.

Appendix.

Since this paper was written and submitted, a survey [23] of the results in this paper has been submitted to the proceedings of a school held in Coimbra, Portugal during May and June of 2001. This survey offers many illustrative examples (with figures). Also a slight improvement has been made in the definition of the standard 2-complex so that cells whose boundaries are labeled by Dyck words are removed. In addition to streamlining the construction (and making sure that it agrees exactly with the group theoretic construction in the case of a group), this allows us to formulate a slightly stronger version of Theorem 8.10.

References


Received June 6, 2001 and revised December 14, 2001. The author was supported in part by NSF-NATO postdoctoral fellowship DGE-9972697 and by FCT through Centro de Matemática da Universidade do Porto.

DEPARTAMENTO DE MATEMÁTICA PURA
FACULDADE DE CIÊNCIAS
DA UNIVERSIDADE DO PORTO
4099-002 PORTO
PORTUGAL

E-mail address: bsteinbg@agc0.fc.up.pt
## CONTENTS

Volume 208, no. 1 and no. 2

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdelmalek Azizi: Construction de la tour des 2-corps de classes de Hilbert de certains corps biquadratiques</td>
<td>1</td>
</tr>
<tr>
<td>Janez Bernik: Brauer-type results on semigroups over p-adic fields</td>
<td>201</td>
</tr>
<tr>
<td>Indranil Biswas and S. Subramanian: Connections on principal bundles over curves in positive characteristics</td>
<td>11</td>
</tr>
<tr>
<td>Holger Brenner and Stefan Schröer: Ample families, multihomogeneous spectra, and algebraization of formal schemes</td>
<td>209</td>
</tr>
<tr>
<td>Heng Huat Chan, Victor Tan and Alice Gee: Cubic singular moduli, Ramanujan’s class invariants $\lambda_n$ and the explicit Shimura Reciprocity Law</td>
<td>23</td>
</tr>
<tr>
<td>Kwang-Wu Chen: Sums of products of generalized Bernoulli polynomials</td>
<td>39</td>
</tr>
<tr>
<td>Francisco-Javier Cirre: The moduli space of real algebraic curves of genus 2</td>
<td>53</td>
</tr>
<tr>
<td>Ulrich Clarenz: A stability criterion for extremals of elliptic parametric functionals</td>
<td>231</td>
</tr>
<tr>
<td>Siddhartha Gadgil: Equivariant framings, lens spaces and contact structures</td>
<td>73</td>
</tr>
<tr>
<td>Alice Gee with Heng Huat Chan and Victor Tan</td>
<td>23</td>
</tr>
<tr>
<td>Eddy Godelle: Parabolic subgroups of Artin groups of type $FC$</td>
<td>243</td>
</tr>
<tr>
<td>Ilseop Han: Hasse principles for the Brauer groups of algebraic function fields of genus zero over global fields</td>
<td>255</td>
</tr>
<tr>
<td>Håkan Hedenmalm and Eero Saksman: Carleson’s convergence theorem for Dirichlet series</td>
<td>85</td>
</tr>
<tr>
<td>Cory Heilmann with Ben W. Reichardt, Yuan Y. Lai and Anita Spielman</td>
<td>347</td>
</tr>
<tr>
<td>Thomas A. Ivey: An Inverse Problem from sub-Riemannian geometry</td>
<td>111</td>
</tr>
<tr>
<td>Yuan Y. Lai with Ben W. Reichardt, Cory Heilmann and Anita Spielman</td>
<td>347</td>
</tr>
<tr>
<td>W.B.R. Lickorish: Knotted contractible 4-manifolds in $S^4$</td>
<td>283</td>
</tr>
<tr>
<td>Aleksandar Mijatović: Simplifying triangulations of $S^3$</td>
<td>291</td>
</tr>
<tr>
<td>Pengcheng Niu and Huiqing Zhang: Payne–Polya–Weinberger type inequalities for eigenvalues of nonelliptic operators</td>
<td>325</td>
</tr>
<tr>
<td>Gianluca Occhetta: Small contractions of smooth varieties</td>
<td>125</td>
</tr>
<tr>
<td>Mainak Poddar: Orbifold Hodge numbers of Calabi–Yau hypersurfaces</td>
<td>151</td>
</tr>
</tbody>
</table>
Ben W. Reichardt, Cory Heilmann, Yuan Y. Lai and Anita Spielman: *Proof of the Double Bubble Conjecture in $\mathbb{R}^4$ and certain higher dimensional cases* 347

Eero Saksman with Håkan Hedenmalm 85

Stefan Schröer with Holger Brenner 209

Anita Spielman with Ben W. Reichardt, Cory Heilmann and Yuan Y. Lai 347

Benjamin Steinberg: *A topological approach to inverse and regular semigroups* 367

S. Subramanian with Indranil Biswas 11

Victor Tan with Heng Huat Chan and Alice Gee 23

Stefano Vidussi: *Norms on the cohomology of a 3-manifold and SW theory* 169

Rachel J. Weir: *Zeros of extremal functions in weighted Bergman spaces* 187

Huiqing Zhang with Pengcheng Niu 325
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 \textsc{La}\textsc{t}e\textsc{x}, but papers in other varieties of \textsc{Te}\textsc{x}, 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 \textsc{La}\textsc{t}e\textsc{x} 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 \textsc{Bib}\textsc{Te}\textsc{x} 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.
Brauer-type results on semigroups over $p$-adic fields

JANEZ BERNIK

Ample families, multihomogeneous spectra, and algebraization of formal schemes

HOLGER BRENNER AND STEFAN SCHROER

A stability criterion for extremals of elliptic parametric functionals

ULRICH CLARENZ

Parabolic subgroups of Artin groups of type FC

EDDY GODDELE

Hasse principles for the Brauer groups of algebraic function fields of genus zero over global fields

ILSEOP HAN

Knotted contractible 4-manifolds in $S^4$

W.B.R. LICKORISH

Simplifying triangulations of $S^1$

ALEKSANDAR MIJATOVIC

Payne–Pólya–Weinberger type inequalities for eigenvalues of nonelliptic operators

PENGCHENG NIU AND HUIQING ZHANG

Proof of the Double Bubble Conjecture in $R^4$ and certain higher dimensional cases

BEN W. REICHARDT, CORY HEILMANN, YUAN Y. LAI AND ANITA SPIELMAN

A topological approach to inverse and regular semigroups

BENJAMIN STEINBERG