Received: 15 February 2021 | Revised: 6 October 2021 | Accepted: 5 November 2021

DOI: 10.1112/jlms.12590

RESEARCH ARTICLE

Journal of the London Mathematical Society

Division in group rings of surface groups

Grigori Avramidi

Max Planck Institute for Mathematics, Bonn, Germany

Correspondence
Grigori Avramidi, Max Planck Institute for Mathematics, Bonn, Germany, 53111.
Email: gavramidi@mpim-bonn.mpg.de

Abstract

We prove a division algorithm for group rings of high genus surface groups and use it to show that some 2-complexes with surface fundamental groups are standard. The division algorithm works somewhat more generally for groups acting on hyperbolic space $\mathbb{H}^n$ with large infimum displacement. We give an application of this to cohomological dimension of 2-relator groups acting on $\mathbb{H}^n$ and to handle decompositions of hyperbolic n-manifolds.

MSC 2020
51F99, 20F67 (primary), 57S30, 16D80 (secondary)

1 | INTRODUCTION

The goal of this paper is to study 2-complexes X with a fixed fundamental group Γ up to homotopy equivalence by means of a division algorithm over the group ring of Γ. These two things are related through the second homotopy group, which is a $\mathbb{Z}\Gamma$-module. Most of the mathematical content of the paper consists of proving a division algorithm for group rings of high genus surface groups. We find this interesting in its own right, even outside the context of 2-complexes.

On division

Let F_n be the free group on n generators. In the rational group ring of a free group there is a division algorithm analogous to polynomial long division that was discovered by Moritz Cohn [5]. A division algorithm is a process that lets one divide one element x by another non-zero element y with a remainder r whose "size" is smaller than that of y. In the group ring $\mathbb{Q}F_n$, the measure of "size" we use is the diameter of the support of the group ring element (defined at the end of this section), which we denote by $|\cdot|$. In symbols, a division algorithm asks for $q, r \in \mathbb{Q}F_n$ such that $x = qy + r$ and $|r| < |y|$ or $r = 0$. Unlike in the case of polynomial long division, there cannot be a division algorithm for non-abelian free groups that works for arbitrary x and y. In fact, for a generic pair of group ring elements, the diameter of the support of any linear combination will be at least as large as that of either element, so there is no hope of obtaining a remainder of smaller diameter. Therefore, in order to have hope there must be linear combinations of x and y of small diameter. What Cohn discovered is that there is a division algorithm *as long as x and y satisfy a non-trivial linear relation in the group ring*. This condition means that there are elements $a, b \in \mathbb{Q}F_n$, not both zero, such that $ax + by = 0$. In fact, a geometric picture of this relation is what dictates the process for actually running the algorithm (see Section 2).

In this paper, we show that the same division algorithm is true when Γ is the fundamental group of a surface of sufficiently high genus. In applications we will also need the division algorithm over the finite fields $\mathbb{F}_p$. So, we state the theorem for a general coefficient field K.

Theorem 1 (Division algorithm for surface groups). *Let K be a field. Let Γ be the fundamental group of a closed, orientable, surface of genus $\geqslant e^{1000000}$. Suppose that x and y are elements in $K\Gamma$ satisfying a non-trivial relation $ax + by = 0$, and $y \neq 0$. Then there are $q, r \in K\Gamma$ such that $x = qy + r$ and $|r| < |y|$ or $r = 0$.*

Our method is inspired by Hog–Angeloni's geometric proof of Cohn's division algorithm [11] and by Delzant's proof that groups rings of hyperbolic groups with large infimum displacement have no zero divisors [8].

Euclid's algorithm for finding the greatest common divisor

The process of applying the division algorithm repeatedly to a pair of elements, dividing at each stage the divisor from the previous stage by the remainder is called Euclid's algorithm. Starting from the division algorithm in the integers (or in the polynomial ring $\mathbb{Q}[t]$) Euclid's algorithm produces the greatest common divisor of two integers (or polynomials). The same is true in our case.

Corollary 2 (Euclid's algorithm for surface groups). *Applying the division algorithm repeatedly, first dividing x by y to obtain a remainder r_1, then dividing y by r_1 to obtain a remainder r_2, and so on, eventually produces an element $z := r_k$ that divides the previous r_{k-1} with no remainder. The element z obtained in this way is a greatest*[†] *common divisor of x and y.*

Algebraic application

Rephrasing things a bit, Euclid's algorithm implies that the (left) ideal (x, y) generated by two elements $x, y \in K\Gamma$ is always free: If x and y do not satisfy any relation, then they are a free basis for the ideal, and if they do satisfy a relation, then the ideal is generated by their greatest common

[†] We say that z is a *divisor* of x if $x = az$ for some $a \in K\Gamma$. It is a *greatest common divisor* of x and y if z is a divisor of x and y and for any other divisor z' of x and y, z' divides z. We say "a" here instead of "the" because greatest common divisors are only well defined up to multiplication by a unit in $K\Gamma$.

divisor z. But, by the theorem of Delzant alluded to earlier, z is not a zero-divisor, which is the same as saying that the ideal z generates is free. Now, let $K\Gamma^d$ be the free $K\Gamma$-module of rank d. We will call elements of this free module *vectors*. A similar argument shows that any two vectors $v, w \in K\Gamma^d$ generate a free $K\Gamma$-module.

Corollary 3. *For any field K, any submodule M of $K\Gamma^d$ generated by two vectors is free.*

For topological applications, we need this sort of result when M is a submodule of the integral group ring $\mathbb{Z}\Gamma$. A "local-to-global" method of Bass ([1]) let one assemble the $\mathbb{Q}$ and $\mathbb{F}_p$ statements together to prove such a result under the additional assumption that the module M is projective (see Corollary 20). We observe that this method applies under the weaker assumption that *the quotient $\mathbb{Z}\Gamma^d/M$ is torsion-free as an abelian group*. This is good enough for us since the topologically meaningful modules associated to a 2-complex satisfy this condition.

Corollary 4. *If a submodule M of $\mathbb{Z}\Gamma^d$ is generated by two vectors and $\mathbb{Z}\Gamma^d/M$ is torsion-free, then M is free.*

Non-free examples

To put the division algorithm into context, note that the statement of Corollary 4 is false for the group $\mathbb{Z}^2$: The ideal $(s-1, t-1)$ in $\mathbb{Z}[\mathbb{Z}^2] = \mathbb{Z}[s, s^{-1}, t, t^{-1}]$ is not free since it has the obvious relation $(s-1)(t-1) = (t-1)(s-1)$ and cannot be generated by one element. More generally, for any non-free group Γ generated two elements a and b, the *augmentation ideal* $(a-1, b-1)$ in $\mathbb{Z}\Gamma$ is not free,[†] so the statement of Corollary 4 is also false for such a group. Such groups arise, by Thurston's work ([16, 4.7]), as fundamental groups of closed hyperbolic 3-manifolds obtained by Dehn filling the figure-eight knot complement (see Section 6). So, the division algorithm and its corollaries do not extend to fundamental groups of arbitrary hyperbolic manifolds.

Groups acting on hyperbolic space with large displacement

Our proof of the division algorithm and its corollaries does work word-for-word for any group Γ that acts by isometries on hyperbolic space $\mathbb{H}^n$ with large *infimum displacement*, which is the infimum of the distances by which non-trivial elements of Γ move points in $\mathbb{H}^n$, that is, $\inf_{x\in\mathbb{H}^n, 1\neq\gamma\in\Gamma} d(x, \gamma x)$.

Theorem 5. *Theorem 1 and Corollaries 2–4 hold for any group Γ that acts on hyperbolic space $\mathbb{H}^n$ with infimum displacement $\geqslant 2000$.*

Applying Corollary 4 to the augmentation ideal of a group satisfying the condition in Theorem 5 shows that any 2-generator group acting on hyperbolic space with large infimum displacement has cohomological dimension one and hence, by Stallings' theorem, is a free group. This recovers (special cases of) freedom theorems of Delzant ([7]) and Gromov ([10, 5.3A]). Applying it to the relation module gives a new theorem about 2-relator groups.

[†] If $(a-1, b-1)$ is free, then Γ has cohomological dimension one, hence is free by Stallings' theorem ([15]).

Corollary 6. *Suppose that Γ is a finitely generated 2-relator group acting by isometries on $\mathbb{H}^n$ with infimum displacement $\geqslant 2000$. Then Γ has cohomological dimension $\leqslant 2$.*

Another easy consequence of the same sort, but of a more geometric flavor, is as follows.

Corollary 7. *Suppose that M is a closed hyperbolic n-manifold of injectivity radius $\geqslant 1000$. Then any handle decomposition for M has at least three k-handles in each dimension $0 < k < n$.*

By contrast, the hyperbolic 3-manifolds mentioned above (obtained by Dehn filling the figure-eight knot complement) only require two k-handles in each dimension.

Topological application

Let us now turn to the topological application mentioned at the beginning of the introduction. An old theorem of Tietze [6] says that two 2-complexes with the same fundamental group and Euler characteristic become homotopy equivalent after wedging both of them with the same sufficiently large number of 2-spheres. A basic question is to determine whether wedging on these extra 2-spheres is really necessary. One of the first examples of inequivalent 2-complexes with the same fundamental group and Euler characteristic involves the trefoil group $T =< a, b \mid a^2 = b^3 >$. Let Y be the presentation 2-complex corresponding to this standard presentation. Dunwoody constructed another presentation 2-complex X for the trefoil group whose second homotopy group $\pi_2 X$ is not free as a $\mathbb{Z}T$-module ([9]). This complex has two generators and two relations, so it has the same Euler characteristic as $Y \vee S^2$, but is not homotopy equivalent to it ($\pi_2(Y \vee S^2)$ is free since Y is aspherical). Dunwoody also showed that the complexes X and $Y \vee S^2$ do become homotopy equivalent after wedging on another S^2, which on the level of π_2 says that $\pi_2 X \oplus \mathbb{Z}T = \mathbb{Z}T \oplus \mathbb{Z}T$. So, $\pi_2 X$ is generated by two elements and is stably free but not free. Corollary 4 implies that this algebraic phenomenon does not happen for fundamental groups Γ of high genus surfaces. (In fact, the weaker Corollary 20 also implies this conclusion.)

We can also ask whether a similar topological phenomenon to the one discovered by Dunwoody can happen for surface groups $\Gamma = \pi_1 \Sigma$ in place of the trefoil group T. If X is a 2-complex with surface fundamental group and minimal Euler characteristic $\chi(X) = \chi(\Sigma)$, then it is easy to see that X is homotopy equivalent to Σ. The first interesting case when the Euler characteristic is non-minimal is $\chi(X) = \chi(\Sigma) + 1$. The main point is to show that $\pi_2 X$ is free. One way is to use a theorem of Louder ([13]) which implies (see Section 7) that X becomes standard after wedging on #(2-cells of X) $- (\chi(X) - \chi(\Sigma))$ different 2-spheres. So, if X has two 2-cells, then $X \vee S^2$ is homotopy equivalent to $\Sigma \vee S^2 \vee S^2$. On π_2, this implies that $\pi_2 X$ is stably free and generated by two elements. If the surface has high enough genus, then Corollary 4 (or 20) implies that $\pi_2 X$ is free, and hence X is homotopy equivalent to $\Sigma \vee S^2$.

Theorem 8. *Let Σ be a closed, orientable surface of genus $\geqslant e^{1000000}$. Suppose that X is a 2-complex with two 2-cells and fundamental group $\pi_1 X = \pi_1 \Sigma$. Then X is homotopy equivalent to Σ or $\Sigma \vee S^2$.*

In fact, a slightly different way to prove the theorem is to use the full strength of Corollary 4 to bypass Louder's result by first showing that the relation module of X (which has no a priori reason to be stably free or projective, but is generated by two relations and does satisfy the hypothesis of Corollary 4) is free and then using this fact to show that $\pi_2(X)$ is generated by two elements and, by another application of Corollary 4, is free as well. This approach has the advantage of applying more broadly to the groups in Corollary 6 (see Section 6).

On 2-complexes with more 2-cells

Let us finish this introduction with several remarks about generalizations to 2-complexes with more than two 2-cells.

For the torus group $\mathbb{Z}^2$ not every submodule of a free $\mathbb{Z}[\mathbb{Z}^2]$-module is free, but all the stably-free ones are (this is Serre's conjecture proved by Quillen and Suslin, see [12]), and this is all one needs to show that any 2-complex with $\mathbb{Z}^2$ fundamental group is standard. For the free groups F_m, there is a generalization of Euclid's algorithm (also due to Cohn) which shows that any ideal in $\mathbb{Q}F_m$ (on any finite number of generators) is free. It also works with coefficients in $\mathbb{F}_p$ instead of $\mathbb{Q}$, so Bass's theorem implies any stably free $\mathbb{Z}F_n$-module is free. This implies that all finite 2-complexes with free fundamental group are standard. (See [11].)

On the other hand, the fundamental group of an orientable genus g surface does has a non-free ideal on $2g$ generators, namely, its augmentation ideal. It is tempting to conjecture that any ideal on fewer than $2g$ generators is free.

Plan of the paper

We explain the division algorithm for free groups in Section 2. In Section 3 we recall and derive properties of hyperbolic space that will be used in the proof of the division algorithm for surface groups (and more generally, groups acting on hyperbolic space with large displacement), which is given in Section 4. We then give a proof of Euclid's algorithm together with Corollaries 3 and 4 in Section 5. The geometric and group-theoretic applications (Corollaries 6 and 7) and one way to get Theorem 8 are proved in Section 6 and the other way is given in Section 7.

Notation and terminology

Before we start, let us fix some notation that will be used throughout the paper and describe how group ring elements can, to a large extent, be thought of geometrically.

Throughout the paper Γ will denote a group acting by covering translations on a space Y, which is either a tree or a hyperbolic space $\mathbb{H}^n$. Pick an orbit of Γ in Y and identify group elements with points of that orbit in Y. Let K be a field. A group ring element $x \in K\Gamma$ is a finite formal linear combination $x = \sum x_\gamma \cdot \gamma$.

Support

The support of x consists of all the group elements γ with non-zero coefficients x_γ appearing in this sum, thought of as points in Y. We will denote the support of an element by the corresponding capital letter. So, the support of x will be denoted as X. We will sometimes refer to points in the support of x as "points of x."

Diameter

The diameter of X is the maximal distance between a pair of points in X. It will be denoted by $|x|$ (or $|X|$), and we will also call it the diameter of x.

Barycenter

The barycenter of X is the center of the smallest closed ball containing X. It will be denoted as $\hat{x}$ (or $\hat{X}$), and we will simply call it the barycenter of x (or X).

Boundary points

Let $B_{\hat{x}}(R)$ be the smallest closed ball containing X. We will call points of X that are a maximal distance R from the barycenter the boundary points of x (or X).

2 | THE DIVISION ALGORITHM FOR FREE GROUPS

In this section, we describe the division algorithm for free groups and sketch its proof.

The algorithm

Suppose that we have a pair of group ring elements x and y that are related by a non-trivial linear relation $ax + by = 0$. The main step in the division algorithm is to show that if $|x| \geqslant |y|$, then we can subtract translates of y from x to obtain an element $x_1 = x - c_1 y$ whose diameter is strictly smaller than that of x. Iterating this step will give division (we will say a few more words about this iteration at the end of this subsection.)

The choice of c_1 is dictated by the relation $ax + by = 0$ as follows. Let o be the barycenter of the support of ax and R the radius of the smallest ball $B_o(R)$ containing this support. There is an x-translate γx with $a_\gamma \neq 0$ whose support γX contains a boundary point of ax. We can assume that $\gamma = 1$, so that X contains a boundary point. (If $\gamma \neq 1$, multiply the relation on the left with γ^{-1} and start again.) Let us call the points of x that are boundary points of ax the *extremal points of x*.

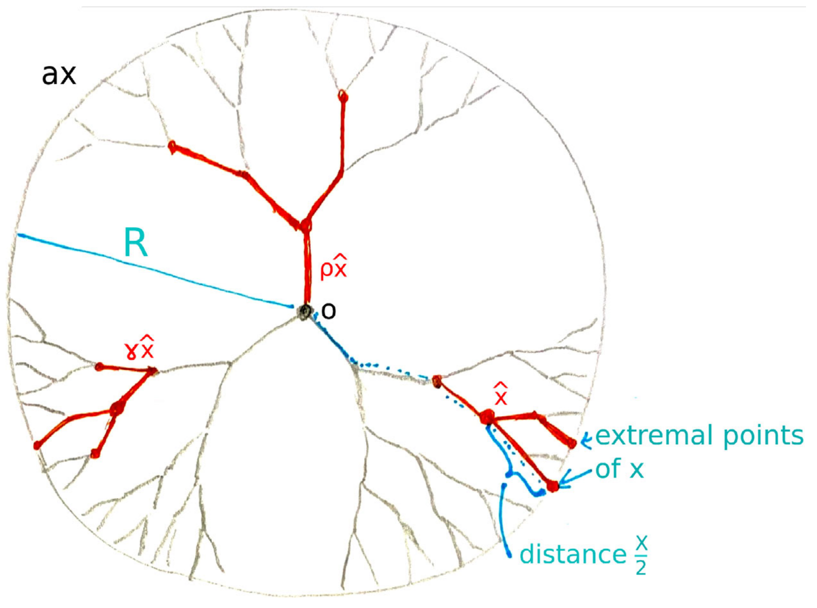
Claim 1. Any boundary point of $ax = -by$ appears in a unique x-translate (that is, γx with $a_\gamma \neq 0$) and also in a unique y-translate (ρy with $b_\rho \neq 0$.)

Therefore, the extremal points of x can all be canceled by y-translates (weighted with appropriate coefficients)[†] to obtain an element

$$x_1 = x - \sum c_\gamma \gamma y,$$

whose support does not contain any of the extremal points from the support of x.

[†] The coefficients are $c_\gamma = -b_\gamma / a_1$ if γy contains an extremal point of x, and $c_\gamma = 0$ otherwise.



Claim 2. x_1 has strictly smaller diameter than x.

If $|x_1| < |y|$, then we take x_1 to be the remainder. If not, then we note that x_1 and y are related by the non-trivial relation $ax_1 + (b + ac_1)y = 0$ and repeat the above argument. Each iteration decreases the diameter by at least one, so after finitely many steps we arrive at an element $x_n = x_{n-1} - c_n y = x - (c_1 + \cdots + c_n)y$ whose diameter is smaller than y. This is our remainder.

Why it works

The key is that we are on a tree. We repeatedly use the following observation.

- If $B_o(R)$ is a closed ball containing a finite set X and $p \in X$ is a point on the boundary $S_o(R)$ of this ball, then the barycenter $\hat{X}$ of X lies on the geodesic from o to p and is precisely $|X|/2$ away from p.

Proof. First, pick a diameter realizing segment qq' with endpoints in X and let m be its midpoint. For any other point $r \in X$, either rmq or rmq' is a geodesic, say the first one. Then $|X| \geqslant d(r,q) = d(r,m) + d(m,q) = d(r,m) + |X|/2$ shows that $d(r,m) \leqslant |X|/2$, that is, the closed ball $B_m(|X|/2)$ contains X. Now, let $B_{\hat{X}}(R_0)$ be the *smallest* closed ball containing X. Since $X \subset B_m(|X|/2)$, we must have $R_0 \leqslant |X|/2$. Either $\hat{X}mq$ or $\hat{X}mq'$ is a geodesic, say the first one. So $|X|/2 \geqslant R_0 \geqslant d(\hat{X},q) = d(\hat{X},m) + d(m,q) = d(\hat{X},m) + |X|/2$, implying that $m = \hat{X}$ and $R_0 = |X|/2$. In summary, the smallest closed ball containing X is $B_{\hat{X}}(|X|/2)$ and any diameter realizing segment of X has $\hat{X}$ as its midpoint.

Now we turn our attention to $B_o(R)$. At least one of $o\hat{X}q$ or $o\hat{X}q'$ is a geodesic, say the first one. Its length is $d(o,q) = d(o,\hat{X}) + |X|/2$. Since X is in $B_o(R)$, we have $R \geqslant d(o,q)$ and so

$$R - |X|/2 \geqslant d(o,\hat{X}).$$

On the other hand, we have $d(\hat{X}, p) \leqslant |X|/2$ since $p \in X \subset B_{\hat{X}}(|X|/2)$ and $d(o, p) = R$ since p is on the boundary of $B_o(R)$, so

$$d(o, \hat{X}) \geqslant d(o, p) - d(p, \hat{X}) \geqslant R - |X|/2.$$

We conclude that all the centered inequalities are equalities, which can only happen if $\hat{X}$ lies on the geodesic from o to p and is precisely $|X|/2$ away from p. □

Recall that the support of x is denoted as X. To prove Claim 1 we look at the set

$$S = \bigcup_{a_\gamma \neq 0} \gamma X.$$

It contains the support of ax but can be bigger if ax has cancellation. Nonetheless, we will show that *the smallest ball containing the support of* ax is also the smallest ball containing S. To that end, let $B_{o'}(R')$ be the smallest ball containing S, and let $S_{o'}(R') = \partial B_{o'}(R')$ be the boundary sphere of this ball. We will show that any point in $S \cap S_{o'}(R')$ is in the support of ax. For this, it is enough to show that any $p \in S \cap S_{o'}(R')$ lies in precisely one X-translate.

Proof. If $a_\gamma \neq 0$ and γX contains a point $p \in S \cap S_{o'}(R')$, then, by the above bullet, the barycenter $\gamma \hat{x}$ lies on the geodesic from o' to p and is precisely $|X|/2$ away from p. If there is another $a_\rho \neq 0$ with ρX containing p, then $\gamma \hat{x} = \rho \hat{x}$ and hence $\gamma = \rho$. So, the translates ρX and γX are the same. □

It follows from this that the points $S \cap S_{o'}(R')$ all appear in the support of ax. Therefore $o = o'$, $R = R'$, what we have called above the "boundary points of ax" are precisely the set $S \cap S_o(R)$, and every boundary point of ax appears in exactly one x-translate. All the same arguments apply to the expression by. This proves the first claim.

Remark. The figure at the beginning of this section illustrates the situation: the smallest ball containing ax entirely contains the supports of all the x-translates $\{\gamma x\}_{a_\gamma \neq 0}$.

To prove the second claim, one uses similar arguments to show that all the points of x_1 (i) are $\leqslant |x|/2$ away from the barycenter $\hat{x}$, and (ii) are not extremal.

Proof of (i) and (ii). If p is a point in x_1, then either p is in x (so (i) $d(\hat{x}, p) \leqslant \frac{|x|}{2}$ and (ii) p is not extremal by construction of x_1) or p is in some y-translate γy in $B_o(R)$ whose support contains an extremal point of x. In the later case, $\hat{x}$ is obtained by going along the geodesic from $\gamma \hat{y}$ to o for a distance $\frac{|x|-|y|}{2}$. Therefore (i) $d(\hat{x}, p) \leqslant d(\hat{x}, \gamma \hat{y}) + d(\hat{y}, p) \leqslant \frac{|x|-|y|}{2} + \frac{|y|}{2} \leqslant \frac{|x|}{2}$ and (ii) if p is extremal then there must be a non-trivial x-translate ρx in $B_o(R)$ whose support contains p. But then $\rho \hat{x}$ is *also* obtained by going along the geodesic from $\gamma \hat{y}$ to o for a distance $\frac{|x|-|y|}{2}$, so $\hat{x} = \rho \hat{x}$ and hence $x = \rho x$, contradicting the fact that ρx is a non-trivial translate of x. So, p cannot be extremal.

Now, note that (i) implies $|x_1| \leqslant |x|$. In the case of equality, there is a diameter realizing segment in x_1 of length $|x_1| = |x|$ whose midpoint is $\hat{x}$. But then, at least one of its endpoints is extremal, which contradicts (ii). So, we must have $|x_1| < |x|$. This proves the second claim.

Where do relations come from?

We can work backward, starting from an element z to produce pairs of elements satisfying successively more complicated relations: $(z, 0) \to (z, az) \to (z + baz, az) \to (z + baz, az + cz + cbaz) \to \dots$. What the division algorithm implies is that any pair satisfying a non-trivial linear relation is obtained by this process.

3 | TREE-LIKE PROPERTIES OF HYPERBOLIC SPACE

Our proof of the division algorithm for surface groups is based on the tree-like properties of hyperbolic space $\mathbb{H}^n$. In this section we recall these properties in a convenient form and derive some specific consequences that will be used in the proof.

3.1 | δ-hyperbolicity

Everything can be easily obtained from the following basic property.

- There is a universal constant δ so that if pq is a segment with midpoint m and o is any point in hyperbolic space, then one of the paths omp or omq cannot be shortened by more than δ. In symbols
$$\max(d(o, p), d(o, q)) \geqslant d(o, m) + \frac{1}{2} d(p, q) - \delta.$$

Remark. In a tree we can take $\delta = 0$ and in hyperbolic space we can take $\delta = 5$.

It is useful to note that one of the angles $\angle_m(o, p)$ or $\angle_m(o, q)$ is obtuse ($\geqslant \pi/2$), and the maximum is achieved for the endpoint corresponding to this obtuse angle. It follows that

- any geodesic segment connecting a sphere $S_o(R)$ to a larger concentric sphere $S_o(R')$ and not intersecting the interior of $B_o(R)$ has length between $|R' - R|$ and $|R' - R| + \delta$.

Proof. Let m be a point on $S_o(R)$ and q a point on $S_o(R')$. The angle $\angle_m(o, q)$ is obtuse, so $d(o, q) \geqslant d(o, m) + d(m, q) - \delta$. Plugging in $d(o, q) = R'$ and $d(o, m) = R$ gives $d(m, q) \leqslant R' - R + \delta$. The other inequality $R' - R \leqslant d(m, q)$ is clear. □

3.2 | Midpoints and barycenters

A consequence of δ-hyperbolicity is that if pq is a length L segment in an R-ball, then its midpoint m is within $R - L/2 + \delta$ of the center of the ball.

Proof. Let o be the center of the R-ball. Then $R \geqslant \max(d(o, p), d(o, q)) \geqslant d(o, m) + L/2 - \delta$. Another consequence is that any set X of diameter D is contained in a $(D/2 + \delta)$-ball. □

Proof. Let p, q be a pair of points realizing the diameter D and let m be their midpoint. If o is any point in X, then $D \geqslant \max(d(o, p), d(o, q)) \geqslant d(o, m) + D/2 - \delta$ implies $d(o, m) \leqslant D/2 + \delta$. In other words, X is contained in the $D/2 + \delta$ ball centered at m. □

These two properties together imply that

- the barycenter of a set is 2δ — close to the midpoint of any segment realizing the diameter.

So we can replace one with the other at the expense of a small error.

Next, suppose that X is a set of diameter D, $\hat{x}$ is its barycenter, and o is a point. Then for any diameter realizing segment pq of X with midpoint m we have

$$\max(d(o,p), d(o,q)) \geqslant d(o,m) + \frac{D}{2} - \delta \tag{1}$$

$$\geqslant d(o,\hat{x}) + \frac{D}{2} - 3\delta. \tag{2}$$

For any point p' in X we have $d(o,p') \leqslant d(o,\hat{x}) + d(\hat{x},p') \leqslant d(o,\hat{x}) + D/2 + \delta$ and therefore

$$d(o,\hat{x}) \geqslant d(o,p') - \frac{D}{2} - \delta. \tag{3}$$

Putting these two inequalities together tells us how far the barycenter $\hat{x}$ is from a point o in terms of the diameter of X and the radius of the smallest ball at o containing X.

Lemma 9. *If $B_o(R)$ is the smallest ball centered at o containing X, then*

$$R - \frac{D}{2} - \delta \leqslant d(o,\hat{x}) \leqslant R - \frac{D}{2} + 3\delta.$$

Proof. Plug $d(o,p') = R$ into (3) and $\max(d(o,p), d(o,q)) \leqslant R$ into (2). □

Shrinking the diameter of X

Another application of these inequalities specifies particular points of X to throw out in order to shrink its diameter.

Lemma 10 (Extremal cancellation). *If $B_o(R)$ is the smallest ball centered at o containing X, then the diameter of $X \cap B_o(R-5\delta)$ is strictly less than the diameter of X.*

Proof. If the diameter of $X \cap B_o(R-5\delta)$ is not smaller, one of its diameter realizing segments pq also realizes the diameter of X. Therefore, plugging (3) into (2) and using $R = d(o,p')$ gives $\max(d(o,p), d(o,q)) \geqslant R - 4\delta$, so at least one of the points p or q is outside the $R - 5\delta$ ball centered at o, which is a contradiction. □

3.3 | Fellow traveling

The next treelike feature of hyperbolic space we need is *fellow traveling*. It says that for a pair of points p and q on the boundary of a ball centered at o, the segments pq and po fellow travel until we reach the midpoint of pq, up to an error 4δ. To express it precisely, it is useful to parametrize geodesics. For a geodesic segment pq we denote by $pq(t)$ the point obtained by traveling from p to q for a time t along the geodesic.

Lemma 11 (Fellow traveling property). *For a pair of points $p, q \in S_o(R)$ and $t \leqslant \frac{d(p,q)}{2}$ we have*

$$d(pq(t), po(t)) \leqslant 4\delta.$$

Proof. Let m be the midpoint of pq, $L = d(o, m)$ and $D = d(m, p)$. Let $p' = po(D)$ be the point obtained by going for a time D from p to o and $p'' = op(L)$ the point obtained by traveling for a time L from o to p. Finally, let m' be the midpoint of the geodesic segment mp'. Now, since the angle $\angle_m(o, p)$ is right, it follows that

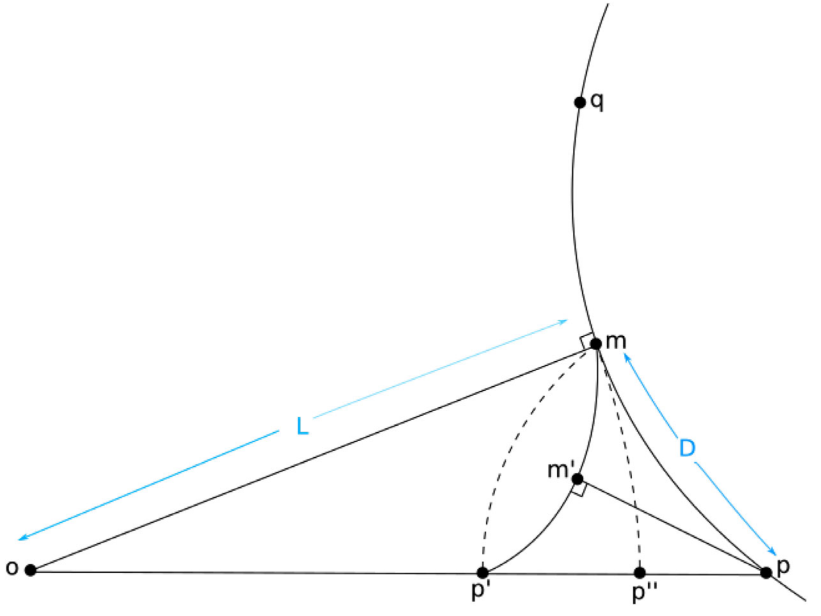
$$R \geqslant L + D - \delta.$$

It is also clear from the picture that

$$d(p, m') \geqslant d(p'', p) = R - L,$$

and plugging in the previous inequality gives

$$d(p, m') \geqslant D - \delta.$$



Since the angle $\angle_{m'}(p, p')$ is right, if follows that

$$D \geqslant d(p, m') + d(m', p') - \delta.$$

Therefore

$$d(m', p') \leqslant D + \delta - d(p, m') \leqslant 2\delta.$$

Since m' is the midpoint of mp', it follows that $d(m, p') \leqslant 4\delta$. This proves the lemma for $t = d(p,q)/2$. The lemma for smaller values of t follows from convexity. $\square$

Large infimum displacement implies no zero divisors in the group ring

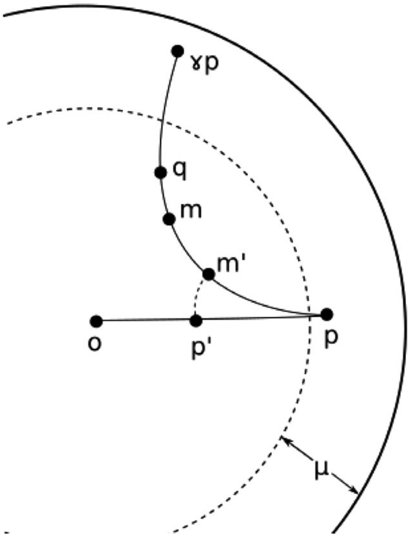
Next, we give a key application of fellow traveling. It is a variation of the main step in Delzant's proof [8] that group rings of groups acting with large displacement on δ-hyperbolic spaces have no zero divisors.

Lemma 12. *Let μ be a non-negative constant. Suppose that γ is an isometry of $\mathbb{H}^n$. If X and γX are contained in a ball $B_o(R)$ and their intersection contains a point p in the μ-neighborhood of the boundary of the ball, then the midpoint m of the segment from p to γp is moved $\leqslant \mu + 9\delta$ by γ^{-1}.*

Proof. Let L be the length of the segment from p to γp. Let q be the point obtained by going from γp to p for a distance $\mu + \delta$. Then $q \in B_o(R - \mu)$. Let m' be the midpoint of the segment from p to q. Then po fellow travels with pq for a time $t = \frac{L-(\mu+\delta)}{2}$ until it reaches m' so, if we denote by $p' = po(t)$ the point reached by traveling from p to o for a time t, then

$$d(m, p') \leqslant d(m, m') + d(m', p') \leqslant \frac{\mu + \delta}{2} + 4\delta.$$

The same argument applied to the segment from p to $\gamma^{-1}p$ shows that its midpoint $\gamma^{-1}m$ satisfies $d(\gamma^{-1}m, p') \leqslant \frac{\mu+\delta}{2} + 4\delta$. Therefore $d(m, \gamma^{-1}m) \leqslant \mu + 9\delta$. □



In other words, if the infimum displacement of Γ acting on $\mathbb{H}^n$ is $> \mu + 9\delta$, then Γ-translates of X that lie in a ball do not intersect in the μ-neighborhood of the boundary of that ball.

Now we apply this to products in the group ring. The following corollary will be used repeatedly in the next section. It implies that any cancellation in a product ax happens away from the boundary of ax, as long as the infimum displacement is sufficiently large.

Corollary 13. *Suppose that Γ has infimum displacement $> \mu + 9\delta$. Let a and x be non-zero group ring elements. Then, the smallest ball containing the support of ax also contains all the X-translates*

$\{\gamma X\}_{a_\gamma \neq 0}$. *Moreover, every point in the μ-neighborhood of the boundary of this ball is contained in at most one such X-translate.*

Proof. The support of the product ax is contained in the set

$$S = \bigcup_{a_\gamma \neq 0} \gamma X.$$

Let $B_o(R)$ be the smallest ball containing S. Lemma 12 implies that on the μ-neighborhood of the boundary of this ball, the X translates $\{\gamma X\}_{a_\gamma \neq 0}$ do not intersect. Therefore, $B_o(R)$ is the smallest ball containing the support of ax. The rest is clear. □

In particular, this says that once the infimum displacement is $> 9\delta$, the support of ax is non-empty, so $ax \neq 0$.

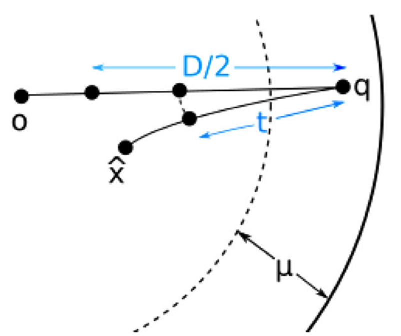
Corollary 14 (Delzant). *Let K be a field. If Γ has infimum displacement $> 9\delta$, then $K\Gamma$ has no zero divisors.*

3.4 | Approximating barycenters

The following lemma is useful.

Lemma 15. *Suppose that X is a set with diameter D and barycenter $\hat{x}$ contained in a ball $B_o(R)$ and $q \in X$ is a point in the μ-neighborhood of the boundary of the ball. Let $qo(D/2)$ be the point obtained by traveling for time $D/2$ along the geodesic from q to o. Then*

$$d(qo(D/2), \hat{x}) \leqslant 9\delta + \frac{3}{2}\mu.$$



Proof. We can assume that $d(q, o) = R - \mu$. First, note that $d(\hat{x}, o) \leqslant R - \frac{D}{2} + 3\delta$ implies that the distance from $\hat{x}$ to $S_o(R-\mu)$ is at least $s = \frac{D}{2} - 3\delta - \mu$. Therefore, the segment $q\hat{x}$ can be extended by s before it reaches $S_o(R-\mu)$, and hence the segment $q\hat{x}$ fellow travels with qo for a time $t = \frac{d(q,\hat{x})+s}{2}$. Also note that the distance from $\hat{x}$ to q is controlled by

$$s \leqslant d(\hat{x}, q) \leqslant \frac{D}{2} + \delta.$$

Therefore

$$|d(q,\hat{x}) - t| = \frac{d(q,\hat{x}) - s}{2} \leqslant \frac{4\delta + \mu}{2},$$

while

$$s \leqslant t \leqslant \frac{D - 2\delta - \mu}{2}$$

implies that

$$|t - D/2| \leqslant 3\delta + \mu.$$

Thus, the distance from $\hat{x}$ to $qo(D/2)$ is bounded by

$$\begin{aligned} d(\hat{x}, qo(D/2)) &\leqslant |d(\hat{x}, q) - t| + 4\delta + |t - D/2| \\ &\leqslant 9\delta + \frac{3}{2}\mu. \end{aligned}$$

□

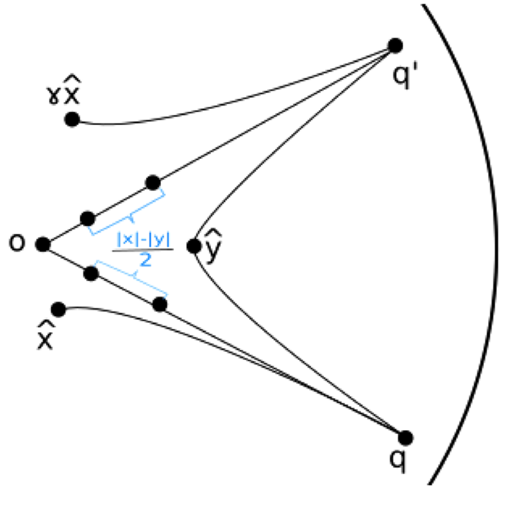
An immediate consequence is the following.

Corollary 16. *If X and Y are both sets in $B_o(R)$ containing a point q in the μ-neighborhood of the boundary, then the barycenters of X and Y are $\frac{||X|-|Y||}{2}$-apart, up to an error $18\delta + 3\mu$.*

A more significant consequence for us is the following.

Corollary 17. *Suppose that $X, \gamma X$, and Y are contained in a ball $B_o(R)$, the intersection of Y and X contains q, and the intersection of Y and γX contains q', where q and q' are in the μ-neighborhood of the boundary of the ball. If $|X| \geqslant |Y|$, then*

$$d(\hat{x}, \gamma\hat{x}) \leqslant 36\delta + 6\mu.$$



Proof. Look at the function $f(t) = d(qo(t), q'o(t))$. It is $\leqslant 18\delta + 3\mu$ at $t = |Y|/2$ by Lemma 15. It decreases as t goes from $|Y|/2$ to $|X|/2$ by convexity (this is where we use $|X| \geqslant |Y|$). Finally, at $t = |X|/2$ it differs from $d(\hat{x}, \gamma\hat{x})$ by an error of at most $18\delta + 3\mu$, again by Lemma 15. This establishes the corollary. □

We will use these two corollaries in the proof of the division algorithm for surface groups in the next section.

4 | PROOF OF THE DIVISION ALGORITHM FOR SURFACE GROUPS

Now, we are ready to prove the division algorithm for group rings of high genus surface groups. The main argument applies more generally to groups Γ acting on hyperbolic space $\mathbb{H}^n$ with large infimum displacement, and yields the following theorem.

Theorem 18 (Division for groups acting with large displacement on hyperbolic space $\mathbb{H}^n$). *Let K be a field. Suppose that Γ acts on $\mathbb{H}^n$ with infimum displacement $\geqslant 2000$. Suppose that x and y are elements in $K\Gamma$ satisfying a non-trivial relation $ax + by = 0$, and $y \neq 0$. Then there are $q, r \in K\Gamma$ such that $x = qy + r$ and $|r| < |y|$ or $r = 0$.*

To obtain the division algorithm for surface groups (Theorem 1) from this, one needs to translate from infimum displacement to genus. This is done at the end of the section.

4.1 | Setup

Throughout the proof, we will keep track of how large the infimum displacement has to be for the argument to work at that stage. To start, we assume that the infimum displacement is $> 9\delta$ so that there are no zero divisors.

We are given a non-trivial relation $ax + by = 0$ where a, b, x and $y \neq 0$ are elements of the group ring $\mathbb{Q}\Gamma$, and we want to show that there are $q, r \in \mathbb{Q}\Gamma$ such that $x = qy + r$ and $|r| < |y|$ or $r = 0$. If $|x| < |y|$, then there is nothing to do, since we can take $q = 0, r = x$. If $|x| \geqslant |y|$, then it is enough to subtract a multiple $b'y$ of y from x for which the resulting element $x' = x - b'y$ has smaller diameter than x. Since the set of possible diameters is discrete and the elements x' and y satisfy the non-trivial[†] relation $a(x - b'y) + (b + ab')y = 0$, iterating the process finitely many times will prove the division algorithm.

Next, we will describe which points of x we will try to cancel out with translates of y in order to reduce the diameter of the resulting group ring element x'. This will be dictated by the relation $ax + by = 0$. Let o be the barycenter of the support of ax and R the radius of ax. As long as the infimum displacement is $> 9\delta$, by Corollary 13, all the x-translates $\{\gamma x\}_{a_\gamma \neq 0}$ are contained in the ball $B_o(R)$. Pick such an x-translate γx containing a boundary point of ax. After multiplying our relation on the left by $(a_\gamma \gamma)^{-1}$, we can assume that this translate is x, that is, that x contains a boundary point of ax and that $a_1 = 1$.

[†] If this relation were trivial, then $a = 0$ and the original relation would imply that y is a zero-divisor.

For a non-negative constant α, let us call the points of x that are in the α-neighborhood of the boundary of ax the *α-extremal* points of x. We have shown in Lemma 10 that if we throw out the 5δ-extremal points from the support of x, then the resulting set has strictly smaller diameter. So these are the points we will try cancel out. To that end, note that if the infimum displacement is $> 5\delta + 9\delta$, then, by Corollary 13, all these 5δ-extremal points are not contained in any other x-translate γx with $a_\gamma \neq 0$. The relation $ax = -by$ implies that each one of them must be contained in a y-translate γy with $b_\gamma \neq 0$, and Corollary 13 applied to by implies that there is a *unique* such y translate.

Therefore, as long as the infimum displacement of Γ is sufficiently large ($> 14\delta$), the 5δ-extremal points of x can all be canceled by y-translates (weighted with appropriate coefficients).[‡] Call the resulting element

$$x' = x - \sum c_\gamma \gamma y.$$

Our goal in the rest of the proof is to show that the diameter of x' is less than the diameter of x.

4.2 | Showing $|x'| < |x|$

Assume that $|x'| \geqslant |x|$. First, we will show that this implies that

- x' contains a 37δ-extremal point.

We begin by finding a ball centered at $\hat{x}$ and containing x'. For a point p of x', if p is a point of x then by Section 3.2, we have

$$d(\hat{x}, p) \leqslant \frac{|x|}{2} + \delta,$$

and if p is a y-point, then by the estimate on the distance between barycenters in Corollary 16:

$$\begin{aligned} d(\hat{x}, p) &\leqslant d(\hat{x}, \gamma\hat{y}) + d(\gamma\hat{y}, p) \\ &\leqslant \left(\frac{|x| - |y|}{2} + 18\delta + 3 \cdot 5\delta\right) + \left(\frac{|y|}{2} + \delta\right) \\ &= \frac{|x|}{2} + 34\delta. \end{aligned}$$

Therefore, the ball of radius $|x|/2 + 34\delta$ centered at $\hat{x}$ contains x'.[†]

Let m be the midpoint of a diameter realizing segment of x'. By Section 3.2, the distance from $\hat{x}$ to m is $\leqslant \frac{|x|}{2} + 34\delta - \frac{|x'|}{2} + \delta \leqslant 35\delta$. Recall that x has a boundary point of ax. Call this point q. Then

$$d(q, m) \leqslant d(q, \hat{x}) + d(\hat{x}, m) \leqslant \frac{|x|}{2} + \delta + 35\delta$$

[‡] To be more precise, $c_\gamma = -b_\gamma$ if γy contains a 5δ-extremal point of x and $c_\gamma = 0$ otherwise.

[†] This implies $|x'| \leqslant |x| + 68\delta$ by the triangle inequality. (Our proof does not need this, but the remark on the next page does.)

shows that m is $(|x|/2 + 36\delta)$-extremal. Since the diameter realizing segment has length $|x'|$, one of its endpoints is $|x|/2 + 36\delta - (|x'|/2 - \delta) \leqslant 37\delta$ extremal. This proves the bullet.

Now, let us look at the extra points $p \in X' - X$ that have been introduced by subtracting the y-translates $\{\gamma y\}_{c_\gamma \neq 0}$ from x. We will refer to points of $X' - X$ as *y-points*, since each of them lies in one of the y-translates $\{\gamma y\}_{c_\gamma \neq 0}$. Fix a constant $\mu \geqslant 37\delta$. We next show that if the infimum displacement is sufficiently large ($> 36\delta + 6\mu$), then such a y-point p cannot be μ-extremal. If it was, then by Corollary 13 it would have to cancel with a unique x-translate ρx that is different from x.[‡] But then the barycenters $\hat{x}$ and $\rho\hat{x}$ would be too close! More precisely, we would have $d(\hat{x}, \rho\hat{x}) \leqslant 36\delta + 6\mu$ by Corollary 17, which contradicts the infimum displacement assumption.

In summary, assuming $|x'| \geqslant |x|$ and infimum displacement $> 36\delta + 6\mu$, we have shown that x' has a 37δ-extremal point of x. Call this point q'.

Remark. In the case of free groups acting on trees, $\delta = 0$ and above we can take $\mu = 0$ so that at this stage in the argument we have an element x' with $|x'| \leqslant |x|$ and if $|x'| = |x|$, then x' has an extremal point, which is not a y-point, hence must be a point of x. But we assumed that all the extremal points of x have been canceled out, so we arrive at a contradiction. In the case of groups acting on $\mathbb{H}^n$, we have to work harder. The reason is that we have found a 37δ-extremal point of x in x', while only the 5δ-extremal points of x have been canceled out.

Now, by Corollary 13, for large enough infimum displacement ($> 37\delta + 9\delta$) the point q' appears in a unique y translate ρy that is different from all the y-translates $\{\gamma y\}_{c_\gamma \neq 0}$ that we subtracted from x to get x'. By Corollary 16 we have

$$d(\hat{x}, \rho\hat{y}) \leqslant \frac{|x| - |y|}{2} + 18\delta + 3 \cdot 37\delta.$$

The rest of the argument breaks up into two cases, depending on the size of $|x| - |y|$.

First, we deal with the case is $|x| - |y| \leqslant \mu$

In this case, the barycenter of ρy and of all the y-translates $\{\gamma y\}_{c_y \neq 0}$ are $(\frac{\mu}{2} + 129\delta)$-close to $\hat{x}$. Therefore, if the infimum displacement is large enough ($> \mu + 258\delta$), we must have $\rho\hat{y} = \gamma\hat{y}$ and hence $\rho = \gamma$, which is a contradiction.

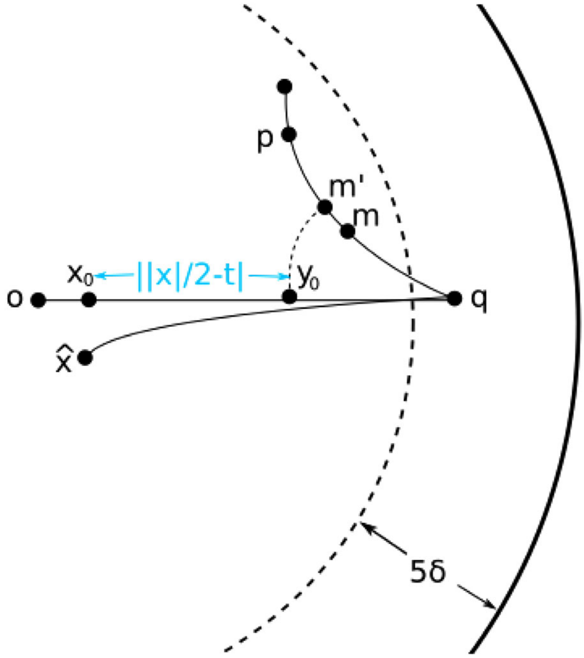
Finally we deal with the case $|x| - |y| \geqslant \mu$

We will show that in this case the y-points of x' are $< |x|/2 - \delta$ away from $\hat{x}$. This will imply that the diameter of x' is less than $|x|$ and we will be done.

Let p be a y-point of x'. Thus, there is a y-translate γy and a 5δ-extremal point q of x so that both p and q are in γy. Denote by L the length of the segment pq and m its midpoint. Since p is not μ-extremal, the segment qp can be extended by $\mu - 5\delta$ before it reaches the 5δ-neighborhood of the boundary. Let m' be the midpoint of this extended segment. It fellow travels with the segment qo for a distance $t = \frac{L+\mu-5\delta}{2}$. Let $y_0 = qo(t)$ be the point obtained by traveling from q to o for a

[‡] If p canceled with x, then it would not have appeared in x'.

time t. Also, let $x_0 = qo(|x|/2)$ be the point obtained by traveling from q to o for a time $|x|/2$. This is illustrated in the figure below.



Note that $\frac{|x|}{2} - t = \frac{|x|-L-(\mu-5\delta)}{2}$ is non-negative because of the case we are in, so we can compute

$$\begin{aligned} d(\hat{x}, p) &\leqslant d(\hat{x}, x_0) + d(x_0, y_0) + d(y_0, m') + d(m', p) \\ &\leqslant \left(9\delta + \frac{3}{2}\cdot 5\delta\right) + \left|\frac{|x|}{2} - t\right| + 4\delta + \frac{L-(\mu-5\delta)}{2} \\ &= \frac{|x|}{2} + 25.5\delta - \mu. \end{aligned}$$

So, for this part of the argument to work, any $\mu > 26.5\delta$ will do.

In summary, everything works for $\mu = 37\delta$, in which case the biggest displacement condition needed is that the infimum displacement is $\mu + 258\delta = 295\delta$. Since we are in hyperbolic space, we can take $\delta = 5$. Since $2000 > 295 \cdot 5$, this finishes the proof of Theorem 18.

4.3 | From infimum displacement to genus

Buser showed in [2] that every surface of genus $\geqslant 2$ has a hyperbolic metric with infimum displacement (=length of shortest geodesic) $\geqslant 2\sqrt{\log(g)}$. For this metric and $g \geqslant e^{1000000}$ we get infimum displacement $\geqslant 2000 > 295 \cdot 5$, which is good enough. This finishes the proof of Theorem 1.

Remark. Buser and Sarnak show in [3] that there is a sequence of hyperbolic surfaces Σ_{g_i} with $g_i \to \infty$ for which one has a much better bound, namely, infimum displacement $\geqslant \frac{4}{3}\log g_i$ and that every genus g surface has a hyperbolic metric with infimum displacement $c \log g$ where c is some small (unspecified) positive constant that does not depend on the genus, but neither of these can be directly applied to get an explicit bound on how high the genus g has to be.

5 | EUCLID'S ALGORITHM AND ALGEBRAIC APPLICATIONS

In this section, Γ is a group acting on hyperbolic space $\mathbb{H}^n$ with infimum displacement $\geqslant 2000$ and K is a field. We will derive Corollaries 2–4 under these assumptions. Together with Theorem 18, this will complete the proof of Theorem 5.

5.1 | Proof of Euclid's algorithm

We are given a pair of elements $x, y \in K\Gamma$ satisfying a non-trivial relation $ax + by = 0$. Dividing x by y we get q_0 and r_0 such that $x = q_0 y + r_0$ and $|r_0| < |y|$ or $r_0 = 0$. If $r_0 \neq 0$, then the elements y and r_0 satisfy the non-trivial relation $ar_0 + (b + aq_0)y = 0$. So we can divide y by r_0 to get q_1 and r_1 such that $y = q_1 r_0 + r_1$ and $|r_1| < |r_0|$ or $r_1 = 0$, and so on. We iterate this process. Since at each step the diameter of the remainder decreases, the process stops after finitely many steps with an r_k that divides r_{k-1} without remainder. All the pairs produced in this way generate the same ideal $(x, y) = (y, r_0) = (r_0, r_1) = \cdots = (r_{k-1}, r_k) = (r_k, 0)$.

We now show that the last remainder $z = r_k$ is a greatest common divisor of x and y. The element z is a divisor of x and y since $x, y \in (z)$. Suppose that z' is another divisor such that $x = cz', y = c'z'$. Since $z \in (x, y)$, we can express it as $K\Gamma$-linear combination $z = a'x + b'y = (a'c + b'c')z'$, so z' is a divisor of z. Therefore, z is a greatest common divisor of x and y.

5.2 | Modules generated by two vectors v, w in $K\Gamma^d$

Delzant's result that $K\Gamma$ has no zero-divisors implies that the submodule of $K\Gamma^d$ generated by a single non-zero vector $v \in K\Gamma^d$ is free. (If there are no zero divisors, then the map $K\Gamma \to K\Gamma^d, a \mapsto av$ is an isomorphism onto its image, which is the module generated by v.) Our division algorithm implies the analogous result for two vectors. The proof is very similar to that of Euclid's algorithm.

Corollary 3. *Let K be a field. Any submodule M of $K\Gamma^d$ generated by two vectors v, w is free.*

Proof. Let $\{v_i\}_{i=1}^d$ and $\{w_i\}_{i=1}^d$ be the coordinates of the vectors v and $w \in K\Gamma^d$, respectively. If both vectors are zero, then there is nothing to do, so we may assume that $v_1 \neq 0$ and that $|v_1| \geqslant |w_1|$. If for any relation $av + bw = 0$ both a and b are zero, then M is free of rank two. So, suppose that there is such a relation with either a or b non-zero. We will show that this implies that M is free of rank one.

There are two cases to consider, depending on whether or not w_1 is zero.

Case 1: $w_1 = 0$. Looking at the first coordinate of the relation, we get $av_1 = 0$ and since $v_1 \neq 0$ we must have $a = 0$. Thus $bw = 0$. Since the relation was non-trivial, $b \neq 0$ so we must have $w = 0$. But then M is generated by a single vector v, and hence it is free of rank one.

Case 2: $w_1 \neq 0$. Then the relation $av_1 + bw_1 = 0$ implies that both a and b are non-zero. We use this relation to divide v_1 by w_1 and get $v' = v - qw$ satisfying $|v_1'| < |w_1|$ or $v_1' = 0$. Then the vectors v', w still generate M and either $v_1' = 0$ or the sum of diameters of their first entries $|v_1'| + |w_1|$ is strictly smaller than $|v_1| + |w_1|$. Moreover, $av' + (b - aq)w = 0$ is again a non-trivial relation (with $a \neq 0$).

At this point, we have arrived back at the situation of the two cases, with v_1' in place of w_1. Moreover, if $v_1' \neq 0$, then the sum of diameters of the first entries of generators $|v_1'| + |w_1|$ is strictly smaller than $|v_1| + |w_1|$. Therefore, after iterating this process finitely many times it will stop and we will arrive in the case 1 situation with M a free module generated by a single vector.

5.3 | Bass's "local-to-global" method for $\mathbb{Z}\Gamma$-modules

Since $\mathbb{Q}\Gamma$ has no zero divisors, its subring $\mathbb{Z}\Gamma$ also has no zero divisors. This implies that any submodule of $\mathbb{Z}\Gamma^d$ generated by a single vector v is free. (If v is zero, there is nothing to show, and if v is non-zero, then the map $\mathbb{Z}\Gamma \to \mathbb{Z}\Gamma^d, a \mapsto av$ is injective, hence an isomorphism onto its image.) The analogous statement for modules generated by two vectors is not true. For example, the ideal $(2, t-1)$ in the group ring $\mathbb{Z}[\mathbb{Z}] = \mathbb{Z}[t, t^{-1}]$ is not free even though the group $\mathbb{Z}$ acts on hyperbolic space with large infimum displacement. A general "local-to-global" theorem of Bass ([1]) shows that this sort of thing does not happen when the module splits off as a direct summand of $\mathbb{Z}\Gamma^d$ (in other words, if the module is *projective*). This is good enough for the proof of Theorem 8 given in Section 7.

Below, we specialize Bass's argument to our situation. For any $\mathbb{Z}\Gamma$-module M, we denote its mod p reduction by $M_p := M/pM$. For example, if v, w are vectors in $\mathbb{Z}\Gamma^d$, then (v, w) is the $\mathbb{Z}\Gamma$-module they generate and $(v, w)_p$ is its mod p reduction.

Lemma 19. *Let v, w be two vectors in $\mathbb{Z}\Gamma^d$. Suppose that the inclusion of modules $(v, w) \hookrightarrow \mathbb{Z}\Gamma^d$ induces inclusions of mod p reductions $(v, w)_p \hookrightarrow \mathbb{F}_p\Gamma^d$ for each prime p. Then (v, w) is a free $\mathbb{Z}\Gamma$-module.*

Proof. If v and w do not satisfy any non-trivial relation, then (v, w) is a free $\mathbb{Z}\Gamma$-module. If they satisfy a non-trivial relation, then they generate the same $\mathbb{Q}\Gamma$ module as their greatest common divisor $z \in \mathbb{Q}\Gamma^d$. We rescale z (multiplying by a rational number if necessary) so that $z \in (v, w)$ and $z \notin (kv, kw)$ for any integer $k > 1$. Since z is a $\mathbb{Q}\Gamma$-divisor of v and w, there is a positive integer m such that $mv = az, mw = bz$ for some $a, b \in \mathbb{Z}\Gamma$. Pick the smallest such m. In summary we have sandwiched the module generated by z in the following way:

$$(mv, mw) \subset (z) \subset (v, w).$$

Our goal is to show that $m = 1$. Suppose that it is not, and let p be a prime dividing m. Note that the composition of induced maps

$$(mv, mw)_p \to (z)_p \to (v, w)_p \tag{4}$$

is zero because p divides m. The key is to show that the second map is injective. □

Claim : The map $i : (z)_p \to (v, w)_p$ is injective. By hypothesis, the inclusion $(v, w) \hookrightarrow \mathbb{Z}\Gamma^d$ induces an inclusion of mod p reductions $(v, w)_p \hookrightarrow \mathbb{F}_p\Gamma^d$. By Corollary 3, $(v, w)_p$ is a free $\mathbb{F}_p\Gamma$-module. So, the image of i is a submodule of a free module and generated by one element, so it is free by Corollary 3. Therefore, i is either injective or the zero map. The later happens precisely if $z \in (pv, pw)$, but this is ruled out by our choice of z. So, the map i is injective.

Since the composition (4) is zero, this implies that the first map $(mv, mw)_p \to (z)_p$ is the zero map, which is the same as saying $(mv, mw) \subset (pz)$. But then z is a $\mathbb{Z}\Gamma$-divisor of both $\frac{m}{p}v$ and $\frac{m}{p}w$, which contradicts the minimality of m. So we are done.

Corollary 20 (Bass). *If a submodule M of $\mathbb{Z}\Gamma^d$ generated by two vectors v, w splits off as a direct summand, then it is free.*

Proof. If (v, w) is a direct summand of $\mathbb{Z}\Gamma^d$, then the composition of the inclusion and projection $(v, w) \hookrightarrow \mathbb{Z}\Gamma^d \to (v, w)$ is the identity map, so its mod p reduction is as well. Therefore, the mod p reduction of the inclusion is injective. So, by Lemma 19, (v, w) is a free module. □

As mentioned in the introduction, one can prove freeness under a weaker assumption on the module M, namely, the assumption that $\mathbb{Z}\Gamma^d/M$ is torsion-free. Note that this assumption is satisfied by the augmentation ideal, the relation module, and the second homotopy module of a 2-complex.

Corollary 4. *If a submodule M of $\mathbb{Z}\Gamma^d$ is generated by two vectors and $\mathbb{Z}\Gamma^d/M$ is torsion-free, then M is free.*

Proof. Suppose $v \in M$ and $v = pw$ for some $w \in \mathbb{Z}\Gamma^d$. In the quotient $Q := \mathbb{Z}\Gamma^d/M$ we have $\overline{v} = 0$ and since Q is torsion-free also $\overline{w} = 0$. But that means $w \in M$ and therefore $v \in pM$. So, we have shown that $M \cap p\mathbb{Z}\Gamma^d = pM$, which is the same as saying that $M_p \to \mathbb{F}_p\Gamma^d$ is injective. So, by Lemma 19, M is a free module. □

6 | GROUPS ACTING ON HYPERBOLIC SPACE

One focus of the present paper is surface groups. They are two-dimensional groups acting on hyperbolic space, and requiring a single relation to present. Moreover, passing to finite index subgroups we get surface groups again but now of higher genus and (if we pick the subgroup correctly) with large infimum displacement. In higher dimensions $n \geqslant 3$ we can start with an arithmetically constructed uniform lattice in $\mathrm{SO}(n, 1)$ and then pass to a deep enough congruence subgroup Γ to get a group action with large infimum displacement on the hyperbolic space $\mathbb{H}^n$. It is well known that — in contrast to surface groups — these higher dimensional lattices require more than one relation. We will show next that they require more than two. This is a direct consequence of the more general statement that any cell structure for $M = \mathbb{H}^n/\Gamma$ has more than two k-cells in each dimension $0 < k < n$.

Corollary 21. *Suppose that M is a closed hyperbolic n-manifold of injectivity radius $\geqslant 1000$. If X is a cell complex homotopy equivalent to M, then X must have at least three k-cells in each dimension $0 < k < n$.*

Proof. Let $\Gamma = \pi_1 M$ be the fundamental group. It acts on $\widetilde{M} = \mathbb{H}^n$ with infimum displacement $\geqslant 2000$. The cellular chain complex of the universal cover of X is a free $\mathbb{Q}\Gamma$-resolution $0 \to C_n \to C_{n-1} \to \cdots \to C_0 \to \mathbb{Q}$, where C_k is a free module of rank equal to the number of k-cells in X. If X has less than three k-cells, then, by Corollary 3, the image of $C_k \to C_{k-1}$ is a free $\mathbb{Q}\Gamma$-module and consequently one gets a free resolution $0 \to$ image $C_k \to C_{k-1} \to \cdots \to C_0 \to \mathbb{Q}$ of length k. But that means that Γ has rational cohomological dimension k, which contradicts the fact that it is the fundamental group of a closed hyperbolic n-manifold. □

Given a handle decomposition for M, one can collapse each k-handle to a k-cell to obtain a homotopy equivalence $M \to X$ to a complex that has one k-cell for each k-handle. Applying the above corollary to this X gives the more geometric sounding Corollary 7 from the introduction.

2-relator groups acting on $\mathbb{H}^n$ with large infimum displacement

Let us now shift attention to 2-relator groups Γ and 2-complexes X presenting them as such. A new wrinkle is that we do not know whether such 2-relator groups have aspherical presentation 2-complexes Y. For any that do (in particular, for the high genus surface groups) it is clear what a *standard* 2-complex with fundamental group Γ is (one homotopy equivalent to $Y \vee S^2 \vee \cdots \vee S^2$). When such a 2-relator group acts on $\mathbb{H}^n$ with large infimum displacement, we get a version of Theorem 8 from the introduction.

Corollary 22. *Suppose that X is a finite 2-complex with two 2-cells and fundamental group Γ. If Γ acts isometrically on $\mathbb{H}^n$ with infimum displacement $\geqslant 2000$, then*

- *the cohomological dimension of Γ is $\leqslant 2$,*
- *$\pi_2 X$ is free, and*
- *if Γ has an aspherical presentation 2-complex Y, then X is homotopy equivalent to either Y or $Y \vee S^2$ or $Y \vee S^2 \vee S^2$. The third case happens only if Γ is a free group.*

Proof. Look at the chain complex on the universal cover:

$$\pi_2(X) \to C_2(\widetilde{X}) \to C_1(\widetilde{X}) \to C_0(\widetilde{X}) \to \mathbb{Z}.$$

The image of the second map is called the relation module, and we will denote it by R. Since C_2 is generated by two elements, its image R is as well. Moreover, R is a submodule of a free module, and the quotient C_1/R is again a submodule of a free module. Therefore, by Corollary 4, R is a free $\mathbb{Z}\Gamma$-module. Since R is also the kernel of the third map, we get a free resolution $R \to C_1 \to C_0 \to \mathbb{Z}$ of length 2. This is the same as saying that the cohomological dimension of Γ is $\leqslant 2$, so we have proved the first bullet.

Since R is free, C_2 splits as a direct sum $\pi_2(X) \oplus R$. Recall that C_2 is generated by two elements, so the projection $C_2 \to \pi_2(X)$ shows that $\pi_2(X)$ is generated by two elements. Moreover, the $\pi_2(X)$ is a submodule of a free module and its quotient $C_2/\pi_2(X)$ is, as well, so Corollary 4 implies that $\pi_2(X)$ is free. This proves the second bullet.

Finally, suppose that there is an aspherical presentation 2-complex Y. Start by building an arbitrary π_1-isomorphism $Y \to X$. Since $\pi_2 X$ is free, we can extend it to a homotopy equivalence from a standard complex Y or $Y \vee S^2$ or $Y \vee S^2 \vee S^2$ by mapping the 2-spheres to a basis for $\pi_2 X$. In the third case $\pi_2 X = \mathbb{Z}\Gamma^2$, so the relation module R vanishes, so Γ has cohomological dimension one and hence, by Stallings' theorem ([15]), is a free group. □

Since the fundamental group of any closed, orientable surface Σ of genus $g \geqslant e^{1000000}$ acts on $\mathbb{H}^2$ with infimum displacement $\geqslant 2000$, the third bullet of Corollary 22 implies Theorem 8.

Flat and hyperbolic 3-dimensional 2-relator groups

Torsion-free 1-relator groups have aspherical presentation 2-complexes ([4]), so they are at most 2-dimensional. We will finish this section with several examples, showing that this is no longer true for 2-relator groups (not even for those 2-relator groups that act by covering translations on hyperbolic space).

The simplest 3-dimensional example of a 2-relator group was pointed out to me by Ian Leary. It is the fundamental group of the mapping torus of the matrix $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ acting on $\mathbb{T}^2$. Note that this is a closed, flat[†] 3-manifold, so the fundamental group is 3-dimensional. It has a 3-generator and 3-relator presentation $< a, b, t \mid [a, b] = 1, tat^{-1} = b, tbt^{-1} = a^{-1} >$. One can eliminate the generator b to get a 2-generator, 2-relator presentation.

There are also hyperbolic 3-manifold examples that were explained to me by Jean Pierre Mutanguha. The mapping torus of the matrix $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ acting on the punctured torus is a hyperbolic 3-manifold with a single cusp.[†] Its presentation is $< a, b, t \mid tat^{-1} = a^2 b, tbt^{-1} = ab >$ and since the second relation says $a = [t, b]$, one can eliminate a together with this relation to get a 1-relator presentation. One can close off the cusp by gluing in a solid torus, and for all but finitely many choices of gluing parameters (a pair of relatively prime numbers (p, q)) one gets a closed hyperbolic 3-manifold (see [16, 4.7]). On the level of fundamental groups, the gluing introduces a new relation of the form $t^p = [a, b]^q$. So, one ends up with a closed hyperbolic 3-manifold whose fundamental group has a 2-generator 2-relator presentation

$$\left\langle b, t \mid t[t, b]t^{-1} = [t, b]^2 b, \quad t^p = [[t, b], b]^q \right\rangle .$$

7 | AN IMPROVED TIETZE'S THEOREM FOR SURFACE FUNDAMENTAL GROUPS

An old theorem of Tietze [6] says that two 2-complexes with the same fundamental group become homotopy equivalent after wedging both of them with enough 2-spheres. This section is about improvements on this theorem when the fundamental group is that of a closed surface Σ. The main point is to interpret a Nielsen equivalence result of Louder in this light.

Minimal Euler characteristic

First note that if X is a 2-complex with fundamental group $\pi_1\Sigma$ and minimal Euler characteristic $\chi(X) = \chi(\Sigma)$, then X is homotopy equivalent to Σ.

[†] The manifold is flat since it is obtained by gluing the ends of $\mathbb{T}^2 \times [0, 1]$ by an isometry.

[†] This manifold is homeomorphic to the figure-eight knot complement (see [17, p. 177]).

Proof. The complexes become homotopy equivalent after wedging both with the same large number of 2-spheres d. Since Σ is aspherical, on π_2 this homotopy equivalence gives $\pi_2 S \oplus \mathbb{Z}\Gamma^d \cong \mathbb{Z}\Gamma^d$. So (see, for example, [14]) S is also aspherical, and hence homotopy equivalent to Σ. □

Nielsen equivalence for surface groups

The orientable surfaces have presentations

$$\left\langle x_1, y_1, \dots, x_g, y_g \mid [x_1, y_1] \cdots [x_g, y_g] = 1 \right\rangle,$$

while the non-orientable ones have presentations

$$\left\langle x_1, \dots, x_r \mid x_1^2 \cdots x_r^2 = 1 \right\rangle.$$

A *standard generating set* is one of these, possibly with some extra generators $z_1, \dots, z_k$ satisfying the trivial relations $z_1 = 1, \dots, z_k = 1$ thrown in at the end.

Now, let X be a finite presentation 2-complex with n generators $e_1, \dots, e_n$ for the surface group, and fix a π_1-isomorphism $f : X \to \Sigma$. In [13], Louder showed the following.

- There is a free group automorphism $\varphi : F_n \to F_n$ so that $f \circ \varphi(e_1), \dots, f \circ \varphi(e_n)$ is a standard generating set for $\pi_1 \Sigma$.

Interpretation as a quantitative variant of Tietze's theorem for surface groups

For concreteness, suppose that it is one representing a genus g orientable surface with k trivial generators at the end (the argument in the non-orientable case is similar). Form a new complex

$$Y = X \cup D_0^2 \cup D_1^2 \cup \cdots \cup D_k^2$$

by attaching $k+1$ different 2-cells to X. The disk D_0^2 is attached along the commutator $[\varphi(e_1), \varphi(e_2)] \cdots [\varphi(e_{2g-1}), \varphi(e_{2g})]$ and the other disks D_i^2 are attached along $\varphi(e_{2g+i})$. By construction, these attaching maps are nullhomotopic in $\pi_1 X$, so Y is homotopy equivalent $X \vee S^2 \vee \cdots \vee S^2$. On the other hand, the map f extends to Y and its restriction to the union $S = D_0^2 \cup \cdots \cup D_k^2$ is a π_1-isomorphism. Since $f : S \to \Sigma$ is a π_1-isomorphism that extends to the 2-cells of X, the attaching maps of the 2-cells of X are null-homotopic in S, and we conclude that Y is also homotopy equivalent to $S^2 \vee \cdots \vee S^2 \vee S$. Finally, since the 2-complex S has the minimal possible Euler characteristic $\chi(S) = \chi(\Sigma)$ among 2-complexes with this fundamental group, the map $f : S \to \Sigma$ is a homotopy equivalence. In summary, X becomes standard after wedging on $k+1$ different 2-spheres:

$$X \vee (k+1)S^2 \sim \Sigma \vee (\#\text{ of 2-cells of } X)S^2.$$

Taking Euler characteristics of this, we see that $k + 1 = (\#\text{ of 2-cells of } X) - (\chi(X) - \chi(\Sigma))$. So, we arrive at the following proposition, which is a quantitative version of Tietze's theorem for surface groups.

Proposition 23. *Let Σ be a closed, orientable surface and X a finite 2-complex with fundamental group $\pi_1\Sigma$. Then X becomes standard after wedging on (# of 2-cells of X) $-$ $(\chi(X) - \chi(\Sigma))$ different 2-spheres.*

Second proof of Theorem 8

The situation that our division algorithm can say something about is when X has one vertex, two 2-cells and Euler characteristic $\chi(X) = \chi(\Sigma) + 1$. In this case, it is easy to see that $k = 0$ and the above homotopy equivalence becomes

$$X \vee S^2 \sim \Sigma \vee S^2 \vee S^2.$$

On π_2 this says $\pi_2 X \oplus \mathbb{Z}\Gamma \cong \mathbb{Z}\Gamma^2$. Therefore, by Corollary 20, $\pi_2 X$ is free. From here we can proceed as in the proof of the third bullet of Corollary 22 to prove Theorem 8. □

ACKNOWLEDGEMENTS

I would like to thank Tâm Nguyễn Phan for suggesting writing Section 2, Ian Leary and Jean Pierre Mutanguha for pointing out some 3-dimensional 2-relator groups, and the Max Planck Institute for its hospitality and financial support.

REFERENCES

1. H. Bass, *Projective modules over free groups are free*, J. Algebra **1** (1964), no. 4, 367–373.
2. P. Buser, *Riemannsche flächen mit grosser kragenweite*, Comment. Math. Helv. **53** (1978), no. 1, 395–407.
3. P. Buser and P. Sarnak, *On the period matrix of a Riemann surface of large genus*, Invent. Math. **117** (1994), no. 1, 27–56.
4. W. H. Cockcroft, *On two-dimensional aspherical complexes*, Proc. Lond. Math. Soc. **3** (1954), no. 1, 375–384.
5. P. M. Cohn, *Free ideal rings and localization in general rings*, vol. 3, Cambridge University Press, Cambridge, 2006.
6. R. H. Crowell and R. H. Fox, *Introduction to knot theory*, vol. 57, Springer Science & Business Media, Cham, 2012.
7. T. Delzant, *Sous-groupes a deux generateurs des groupes hyperboliques*, Group theory from a geometrical viewpoint, World Scientific, Singapore, 1990, pp. 177–189.
8. T. Delzant, *Sur l'anneau d'un groupe hyperbolique*, C. R. Acad. Sci. **324** (1997), no. 4, 381–384.
9. M. J. Dunwoody, *The homotopy type of a two-dimensional complex*, Bull. Lond. Math. Soc. **8** (1976), no. 3, 282–285.
10. M. Gromov, *Hyperbolic groups*, Essays in group theory, Springer, New York, 1987, pp. 75–263.
11. C. Hog-Angeloni, *A short topological proof of Cohn's theorem*, Topology and combinatorial group theory, Springer, New York, 1990, pp. 90–95.
12. T.-Y. Lam, *Serre's problem on projective modules*, Springer Science & Business Media, Berlin, Heidelberg, 2010.
13. L. Louder, *Nielsen equivalence in closed surface groups*, arXiv:1009.0454, 2010.

14. M. S. Montgomery, *Left and right inverses in group algebras*, Bull. Amer. Math. Soc. **75** (1969), no. 3, 539–540.
15. J. Stallings, *On torsion-free groups with infinitely many ends*, Ann. of Math. **88** (1968), 312–334.
16. W. P. Thurston, *The geometry and topology of 3-manifolds*, Lecture note, 1978.
17. W. P. Thurston, *Three dimensional manifolds, Kleinian groups and hyperbolic geometry*, Bull. Lond. Math. Soc. **6** (1982), no. 3, 357–381.