Sharp phase transition theorems for hyperbolicity of random groups

Yann Ollivier

7th February 2003

With 28 illustrations

Abstract

We prove that in various natural models of a random quotient of a group, depending on a density parameter, for each hyperbolic group there is some critical density under which a random quotient is still hyperbolic with high probability, whereas above this critical value a random quotient is very probably trivial. We give explicit characterizations of these critical densities for the various models.

Introduction

What does a generic group look like?

The study of random groups emerged from an affirmation of M. Gromov that "almost every group is hyperbolic" (see [Gro1]). The first proof of such a kind of theorem was given by A.Y. Ol'shanskii in [Ol], and independently by C. Champetier in [Ch1]: fix m and N and consider the group G presented by $\langle a_1, \ldots, a_m; r_1, \ldots r_N\rangle$ where the r_i's are words of length ℓ_i in the letters $a_i^{\pm 1}$. Then the ratio of the number of n-tuples of words r_i such that G is hyperbolic, to the total number of n-tuples of words r_i, tends to 1 as $\inf \ell_i \to \infty$, thus confirming Gromov's statement.

Later, M. Gromov introduced (cf. [Gro2]) a finer model of random group, in which the number N of relators is allowed to be much bigger.

This model goes as follows: Choose at random N reduced words of length ℓ in the letters $a_i^{\pm 1}$, uniformly among the set of all such reduced words (recall a word is called *reduced* if it does not contain a sequence of the form $a_i a_i^{-1}$ or $a_i^{-1} a_i$). Let R be the (random) set of these N words, the random group is defined by presentation $\langle a_1, \ldots, a_m; R\rangle$.

Let us explain how N is taken in this model. There are $(2m)(2m-1)^{\ell-1} \approx (2m-1)^\ell$ reduced words of length ℓ. We thus take $N = (2m-1)^{d\ell}$ for some number d between 0 and 1 called *density*.

The theorem stated by Gromov in this context expresses a sharp phase transition between hyperbolicity and triviality, depending on the asymptotics of the number of relators taken, which is controlled by the density parameter d.

Theorem 1 (M. Gromov, [Gro2]) – *Fix a density d between 0 and 1. Choose a length ℓ and pick at random a set R of $(2m-1)^{d\ell}$ uniformly chosen reduced words of length ℓ in the letters $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$.*

If $d < 1/2$ then the probability that the presentation $\langle a_1, \ldots, a_m; R\rangle$ defines an infinite hyperbolic group tends to 1 as $\ell \to \infty$.

If $d > 1/2$ then the probability that the presentation $\langle a_1, \ldots, a_m; R\rangle$ defines the group $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$ tends to 1 as $\ell \to \infty$.

A complete proof of this theorem in included below (section 2).

Let us discuss the intuition behind this model. What does the density parameter d mean? Following the excellent exposition of Gromov in [Gro2], we assimilate $d\ell$ to a dimension. That is, $d\ell$ represents the number of "equations" we can impose on a random word so that we still have a reasonable chance to find such a word in a set of $(2m-1)^{d\ell}$ randomly chosen words (compare to the basic intersection theory for random sets stated in section 5.2).

For example, for large ℓ, in a set of $2^{d\ell}$ randomly chosen words of length ℓ in the two letters "a" and "b", there will probably be some word beginning with $d\ell$ letters "a". (This is a simple exercise.)

As another example, in a set of $(2m-1)^{d\ell}$ randomly chosen words on $a_i^{\pm 1}$, there will probably be two words having the same first $2d\ell$ letters, but no more. In particular, if $d < 1/12$ then the set of words will satisfy the small cancellation property $C'(1/6)$ (see [GH] for definitions). But as soon as $d > 1/12$, we are far from small cancellation, and as d approaches $1/2$ we have arbitrarily big cancellation.

The purpose of this work is to give similar theorems in a more general situation. The theorem above states that a random quotient of the free group F_m is hyperbolic. A natural question is: does a random quotient of a hyperbolic group stay hyperbolic?

This would allow in particular to iterate the operation of taking a random quotient. This kind of construction is at the heart of the "wild" group constructed in [Gro4].

Our theorems precisely state that for each hyperbolic group, there is a critical density d under which the quotient stays hyperbolic, and above which it is probably trivial.

Moreover, this critical density can be characterized in terms of well-known numerical quantities depending on the group.

There are several ways to generalize Gromov's theorem: a good replacement in a hyperbolic group for reduced words of length ℓ in a free group could, equally likely, either be reduced words of length ℓ again, or elements of norm ℓ in the group (the norm of an element is the minimal length of a word equal to it). We have a theorem for each of these two cases. We also have a theorem for random quotients by uniformly chosen plain words (without any assumption).

In the first two versions, in order to have the number of reduced, or geodesic, words of length ℓ tend to infinity with ℓ, we have to suppose that G is not elementary. There is no problem with the case of a quotient of an elementary group by plain random words (and the critical density is 0 in this case).

Let us begin with the case of reduced words.

We recall the definition and basic properties of the cogrowth η of a group G in section 1.2 below. Basically, if G is not free, the number of reduced words of length ℓ which are equal to e in G behaves like $(2m-1)^{\eta\ell}$. For a free group, η is (conventionally, by the way) equal to $1/2$. It is always at least $1/2$.

Theorem 2 (Random quotient by reduced words) – *Let G be a non-elementary hyperbolic group generated by the elements $a_1, \ldots, a_m$. Fix a density d between 0 and 1. Choose a length ℓ and pick at random a set R of $(2m-1)^{d\ell}$ uniformly chosen reduced words of length ℓ in $a_i^{\pm 1}$. Let $\langle R \rangle$ be the normal subgroup generated by R.*

Let η be the cogrowth of the group G.

If $d < 1 - \eta$, then, with probability tending to 1 as $\ell \to \infty$, the quotient $G/\langle R \rangle$ is non-elementary hyperbolic.

If $d > 1 - \eta$, then, with probability tending to 1 as $\ell \to \infty$, the quotient $G/\langle R \rangle$ is either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$.

We go on with the case of elements on the ℓ-sphere of the group.

In this case, for the triviality part of the theorem, some small-scale phenomena occur, comparable to the occurrence of $\mathbb{Z}/2\mathbb{Z}$ above (think of a random quotient of $\mathbb{Z}$ by any number of elements of norm ℓ). In order to avoid them, we take words of norm not exactly ℓ, but of norm between $\ell - L$ and $\ell + L$ for some fixed $L > 0$ ($L = 1$ is enough).

Theorem 3 (Random quotient by elements of a sphere) – *Let G be a non-elementary hyperbolic group generated by the elements $a_1, \ldots, a_m$. Fix a density d between 0 and 1. Choose a length ℓ.*

Let S^ℓ be the set of elements of G which are of norm between $\ell - L$ and $\ell + L$ with respect to $a_1, \ldots, a_m$ (for some fixed $L > 0$). Let N be the number of elements of S^ℓ.

Pick at random a set R of N^d uniformly chosen elements of S^ℓ. Let $\langle R \rangle$ be the normal subgroup generated by R.

If $d < 1/2$, then, with probability tending to 1 as $\ell \to \infty$, the quotient $G/\langle R \rangle$ is non-elementary hyperbolic.

If $d > 1/2$, then, with probability tending to 1 as $\ell \to \infty$, the quotient $G/\langle R \rangle$ is $\{e\}$.

The two theorems above were two possible generalizations of Gromov's theorem. On can wonder what happens if we completely relax the assumptions on the words, and take in our set R any kind of words of length ℓ with respect to the generating set. The same kind of theorem still applies, with of course a smaller critical density.

The gross cogrowth θ of a group is defined in section 1.2 below. Basically, $1-\theta$ is the exponent (in base $2m$) of return to e of the random walk on the group. We always have $\theta > 1/2$.

Now there are $(2m)^\ell$ candidate words of length ℓ, so we define density with respect to this number.

Theorem 4 (Random quotient by plain words) – *Let G be a hyperbolic group generated by the elements $a_1, \dots, a_m$. Fix a density d between 0 and 1. Choose a length ℓ and pick at random a set R of $(2m)^{d\ell}$ uniformly chosen words of length ℓ in $a_i^{\pm 1}$. Let $\langle R \rangle$ be the normal subgroup generated by R.*

Let θ be the gross cogrowth of the group G.

If $d < 1-\theta$, then, with probability tending to 1 as $\ell \to \infty$, the quotient $G/\langle R \rangle$ is non-elementary hyperbolic.

If $d > 1-\theta$, then, with probability tending to 1 as $\ell \to \infty$, the quotient $G/\langle R \rangle$ is either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$.

Precisions on the models. Several points in the theorems above are left for interpretation.

There is a slight difference between choosing N times a random word and having a random set of N words, since some word could be chosen several times. But for $d < 1/2$ the probability that a word is chosen twice is very small and the difference is negligible; anyway this does not affect our statements at all, so both interpretations are valid.

Numbers such as $(2m)^{d\ell}$ are not necessarily integers. We can either take the integer part, or choose two constants C_1 and C_2 and consider taking the number of words between $C_1(2m)^{d\ell}$ and $C_2(2m)^{d\ell}$. Once more this does not affect our statements at all.

The case $d = 0$ is peculiar since nothing tends to infinity. Say that a random set of density 0 is a random set with a number of elements growing subexponentially in ℓ (e.g. with a constant number of elements).

The possible occurrence of $\mathbb{Z}/2\mathbb{Z}$ above the critical density only reflects the fact that it may be the case that a presentation of G has no relators of odd length (as in the free group). So, when quotienting by words of even length, at least $\mathbb{Z}/2\mathbb{Z}$ remains.

Discussion of the models. Of course, the three theorems given above are not proved separately, but are particular cases of a more general (and more technical!) theorem. This theorem is stated in section 4.4.

Our general theorem deals with random quotients by words picked under a given probability measure. This measure does not need to be uniform, neither does it necessarily charge words of only one given length. It has to satisfy some natural (once the right terminology is given...) axioms. The axioms are stated in section 4.3, and the quite sophisticated terminology for them is given in section 4.2.

For example, using these axioms it is easy to check that taking a random quotient by reduced words or by cyclically reduced words is (asymptotically) the same, with the same critical density.

It is also possible to take quotients by words of different lengths, but our method imposes that the ratio of the lengths be bounded. This is a restriction due to the geometric nature of some parts of the argument, which rely on the hyperbolic local-to-global principle, using metric properties of the Cayley complex of the group (cf. appendix A).

In the case of various lengths, density has to be defined as the supremum of the densities at each length.

The very first model of random group given in this article (the one used by Ol'shanskii and Champetier), with a constant number of words of prescribed lengths, is morally the case $d = 0$ of our models, but not technically, as in this model the ratio of lengths can be unbounded, thus preventing the use of some geometric methods.

But another model encountered in the literature, which consists in uniformly picking a fixed number of words of length between 1 and ℓ, satisfies easily our axioms, as it is almost exactly our case $d = 0$. Indeed there are so much more words of length close to ℓ than close to 0, that the elements taken under this model are of length comprised between $(1-\varepsilon)\ell$ and ℓ for any ε.

Whereas random plain words or random reduced words can be easily constructed independently of the group, it could seem difficult, at first glance, to take a quotient by random elements of a sphere. Let us simply recall (cf. [GH]) that in a hyperbolic group, it is possible to define for each element a normal geodesic form, and that there exists a finite automaton which recognizes exactly the words which are normal forms of elements of the group.

Note that all our models of random quotients depend on a generating subset. For example, adding "false generators" (i.e. generators equal to e) to our generating sets makes the cogrowth and gross cogrowth arbitrarily close to 1, thus the critical density for reduced words and plain words arbitrarily small. The case of random quotients by elements of the ℓ-sphere seems to be more robust.

Other developments on generic properties of groups. Other properties of generic groups have been studied under one or another model of random group. Besides hyperbolicity, this includes topics such as small cancellation properties, torsion elements, topology of the boundary, property T, or the fact that most subgroups are free; and more are to come. See for example [Ch1], [AO], [A], [Z].

Random groups have also been used by M. Gromov to construct a “wild” group related to $C^\star$-algebraic conjectures, see [Gro4].

In a slightly different approach, the study of what a generic group looks like has very interesting recent developments: genericity can also be understood as a topological (rather than probabilistic) property in the space of all finite type groups. See for example the work of C. Champetier in [Ch3].

In all these works, properties linked to hyperbolicity are ubiquitous.

Acknowledgements. I would like to thank, in alphabetical order, Thomas Delzant, Étienne Ghys, Misha Gromov, Claire Kenyon, Richard Kenyon, Pierre Pansu, Panos Papasoglu, Frédéric Paulin and Andrzej Żuk for instructive talks and comments.

Part of the ideas of this work emerged during my stay at the École normale supérieure of Lyon in April 2002, at the invitation of Andrzej Żuk. I would like to thank all the team of the Mathematics Department there for their great warmth at receiving me.

1 Definitions and notations

1.1 Basics

Throughout all this text, G will be a discrete hyperbolic group given by a presentation $\langle S; R\rangle$ where $S = \{a_1, \dots, a_m, a_1^{-1}, \dots a_m^{-1}\}$ is a symmetric set of $2m$ generators, and R is a finite set of words on S. (Every discrete hyperbolic group is finitely presented, cf. [S].)

We shall denote by δ a hyperbolicity constant for G w.r.t. S. Let λ be the maximal length of relations in R.

A hyperbolic group is called *non-elementary* if it is neither finite nor quasi-isometric to $\mathbb{Z}$.

A *word* will be a word made of letters in S. Equality of words will always mean equality as elements of the group G.

A word is said to be *reduced* if it does not contain a generator $a \in S$ immediatly followed by its inverse a^{-1}. It is said to be *cyclically reduced* if it and all of its cyclic permutations are reduced.

If w is a word, we shall call its number of letters its *length* and denote it by $|w|$. Its *norm*, denoted by $\|w\|$, will be the smallest length of a word equal to w in the group G.

1.2 Growth, cogrowth, and gross cogrowth

First, we recall the definition of the growth, cogrowth and gross cogrowth of the group G with respect to the generating set S.

Let S^ℓ be the set of all words of length ℓ in $a_i^{\pm 1}$. Let S^ℓ_G be the set of all elements of G the norm of which is equal to ℓ with respect to the generating set $a_i^{\pm 1}$. The growth controls the asymptotics of the number of elements of S^ℓ_G: this number is roughly equal to $(2m)^{g\ell}$. The gross cogrowth controls the asymptotics of the number of words in S^ℓ which are equal to the neutral element in G: this number is roughly equal to $(2m)^{\theta\ell}$. The cogrowth is the same with reduced words only: this number is roughly $(2m-1)^{\eta\ell}$.

These quantities have been extensively studied. Growth now belongs to the folklore of discrete group theory. Cogrowth has been introduced by R. Grigorchuk in [Gri], and independently by J. Cohen in [C]. For some examples see [Ch2] or [W1]. Gross cogrowth is linked (see below) to the spectrum of the random walk on the group, which, since the seminal work by H. Kesten (see [K1] and [K2]), has been extensively studied (see for example the numerous technical results in [W2] and the references therein).

Definition 5 (Growth, cogrowth, gross cogrowth) –
The growth *of the group G with respect to the generating set $a_1, \dots, a_m$ is defined as*

$$g = \lim_{\ell\to\infty} \frac{1}{\ell} \log_{2m} \# S^\ell_G$$

The gross cogrowth *of the group G with respect to the generating set $a_1, \dots, a_m$ is defined as*

$$\theta = \lim_{\substack{\ell\to\infty \\ \ell \text{ even}}} \frac{1}{\ell} \log_{2m} \# \{ w \in S^\ell, w = e \text{ in } G \}$$

The cogrowth *of the group G with respect to the generating set $a_1, \dots, a_m$ is defined as $\eta = 1/2$ for a free group, and otherwise*

$$\eta = \lim_{\substack{\ell\to\infty \\ \ell \text{ even}}} \frac{1}{\ell} \log_{2m-1} \# \{ w \in S^\ell, w = e \text{ in } G, w \text{ reduced} \}$$

Let us state some properties of these quantities. There are all proved in [K2], [Gri] or [C].

The limits are well-defined by a simple subadditivity (for growth) or superadditivity (for the cogrowths) argument. We restrict ourselves to even ℓ because there may be no word of odd length equal to the trivial element, as is the case e.g. in a free group.

For cogrowth, the logarithm is taken in base $2m-1$ because the number of reduced words of length ℓ behaves like $(2m-1)^\ell$.

The cogrowth and gross cogrowth lie between $1/2$ and 1. Gross cogrowth is strictly above $1/2$, as well as cogrowth except for the free group. There exist groups with cogrowth or gross cogrowth arbitrarily close to $1/2$.

The probability that a random walk in the group G (with respect to the same set of generators) starting at e, comes back to e at time ℓ is equal to the number of words equal to e in G, divided by the total number of words of length ℓ. This leads to the following characterization of gross cogrowth, which says that the return probability at time t is roughly equal to $(2m)^{-(1-\theta)t}$. This will be ubiquitous in our text.

Alternate definition of gross cogrowth – *Let P_t be the probability that a random walk on the group G (with respect to the generating set $a_1, \ldots, a_m$) starting at e at time 0, comes back to e at time t.*

Then the gross cogrowth of G w.r.t. this generating set is equal to

$$\theta = 1 + \lim_{\substack{t \to \infty \\ t \text{ even}}} \frac{1}{t} \log_{2m} P_t$$

In particular, $(2m)^{\theta-1}$ is the spectral radius of the random walk operator (denoted λ in [K1] and r in [Gri]), which is the form under which it is studied in these papers.

A cogrowth, or gross cogrowth, of 1 is equivalent to amenability.

It is easy to check that $g + \theta \geqslant 1$.

Gross cogrowth and cogrowth are linked by the following equation:

$$(2m)^\theta = (2m-1)^\eta + (2m-1)^{1-\eta}$$

The gross cogrowth of the free group F_m is $\frac{1}{2}\log_{2m}(8m-4)$, and this is the only group on m generators with this gross cogrowth. This tends to $1/2$ as $m \to \infty$.

There are various conventions for the cogrowth of the free group. The definition above would give $-\infty$. In [C] the cogrowth of the free group is taken equal to 0; in [Gri] it is not defined. Our convention allows the relation above between cogrowth and gross cogrowth to be valid even for the free group; it is also natural given the fact that, for any group except the free group, the cogrowth is strictly above $1/2$. Moreover, this leads to a single formulation for our random quotient theorem, as with this convention, the critical density for quotients by reduced words will be equal to $1-\eta$ in any case. So we strongly plead for this being the right convention.

If G is presented as F_m/N where N is a normal subgroup, the cogrowth is the growth (in base $2m-1$) of N. The gross cogrowth is the same considering N as a submonoid in the free monoid on $2m$ generators and in base $2m$.

Let Δ be the Laplacian on G (w.r.t. the same generating set). As the operator of convolution by a random walk is equal to $1-\Delta$, we get another characterization of gross cogrowth. The eigenvalues lie in the interval $[0;2]$. Let λ_0 be the smallest one and λ'_0 the largest one. Then the gross cogrowth of G w.r.t. this generating set is equal to

$$\theta = 1 + \log_{2m} \sup(1-\lambda_0, \lambda'_0 - 1)$$

(We have to consider λ'_0 due to parity problems.)

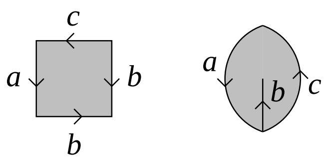
The cogrowth and gross cogrowth depend on the generating set. For example, adding trivial generators $a_i = e$ makes them arbitrarily close to 1.

1.3 Diagrams

A *filamenteous Dehn diagram* in the group G with respect to the presentation $\langle S; R\rangle$ will be a planar connected combinatorial 2-complex decorated in the following way:

- Each 2-cell c bears some relator $r \in R$. The number of edges of the border of c is equal to $|r|$.
- If e is an (unoriented) edge, denote by e_+ and e_- its two orientations. Then e_+ and e_- both bear some generator $a \in S$, and these two generators are inverse.
- Each 2-cell c has a marked vertex on its boundary, and an orientation at this vertex.
- The word read by going through the (oriented) edges of the boundary of cell c, starting at the marked point and in the direction given by the orientation, is the relator $r \in R$ attached to c.

Note on the definition of regular complexes: we do not require that each closed 2-cell be homeomorphic to the standard disc. We only require the interior of the 2-cell to be homeomorphic to a disc, that is, the application may be non-injective on the boundary. This makes a difference only when the relators are not reduced words. For example, if $abb^{-1}c$ is a relator, then the two diagrams below are valid. We will talk about *regular diagrams* to exclude the latter.



We will use the terms 2-*cell* and *face* interchangeably.

A *non-filamenteous* Dehn diagram will be a diagram in which every 1- or 0-cell lies in the boundary of one 2-cell. Unless otherwise stated, in our text *a Dehn diagram will implicitly be non-filamenteous.*

A *n-hole* Dehn diagram will be one for which the underlying 2-complex has n holes. When the number of holes is not given, *a Dehn diagram will be supposed to be simply connected* (0-hole).

A *decorated abstract Dehn diagram* (daDd for short) is defined almost the same way as a Dehn diagram, except that no relators are attached to the 2-cells and no generators attached to the edges, but instead, to each 2-cell is attached an integer between 1 an the number of 2-cells of the diagram (and yet, a starting point and orientation to each 2-cell).

Please note that this definition is a little bit emended in section 6.2 (more decoration is added).

A daDd is said to be *fulfillable* w.r.t. presentation $\langle S; R\rangle$ if there exists an assignment of relators to 2-cells and of generators to 1-cells, such that any two 2-cells bearing the same number get the same relator, and such that the resulting decorated diagram is a Dehn diagram with respect to presentation $\langle S; R\rangle$.

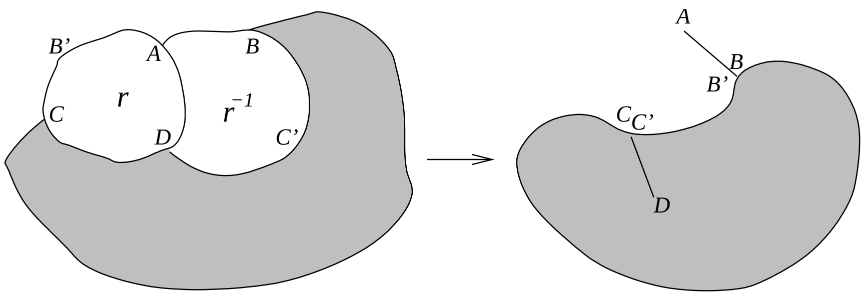
A *daDd with border* $w_1, \dots, w_n$, where $w_1, \dots, w_n$ are words, will be a $(n-1)$-hole daDd with each boundary edge decorated by a letter such that the words read on the n components of the boundary are $w_1, \dots, w_n$. A daDd with border is said to be *fulfillable* if, as a daDd, it is fulfillable while keeping the same boundary words.

A word w is equal to the neutral element e in G if and only if some no-hole daDd with border w is fulfillable (see [LS]).

A Dehn diagram is said to be *reduced* if there is no pair of adjacent (by an edge) 2-cells bearing the same relator with opposite orientations and with the common edge representing the same letter of the relator (w.r.t. the starting point). A daDd is said to be *reduced* if there is no pair of adjacent (by an edge) 2-cells bearing the same number, with opposite orientations and a common edge representing the same letter of the relator.

A Dehn diagram is said to be *minimal* if it has the minimal number of 2-cells among those Dehn diagrams having the same boundary word (or boundary words if it is not simply connected). A fulfillable daDd with border is said to be *minimal* in the same circumstances.

Note that a minimal Dehn diagram is necessarily reduced: if there were a pair of adjacent faces with the same relator in opposite orientations, then they could be removed to obtain a new diagram with less faces and the same boundary (maybe adding some filaments):



Throughout the text, we shall use the term *diagram* as a short-hand for "Dehn diagram or fulfillable decorated abstract Dehn diagram". We will use the term *minimal diagram* as a short-hand for "minimal Dehn diagram or minimal fulfillable decorated abstract Dehn diagram with border".

1.4 Isoperimetry and narrowness

There is a canonical metric on the 1-skeleton of a Dehn diagram (or a daDd), which assigns length 1 to every edge. If D is a diagram, we will denote its number of faces by $|D|$ and the length of its boundary by $|\partial D|$.

It is well-known (see [S]) that a discrete group is hyperbolic if and only if there exists a constant C such that any minimal diagram D satisfies the linear isoperimetric inequality $|\partial D| \geqslant C\,|D|$. We show in Appendix B that in a hyperbolic group, holed diagrams satisfy an isoperimetric inequality as well.

Throughout all the text, C will be an isoperimetric constant for G.

The set of 2-cells of a diagram is also canonically equipped with a metric: two 2-cells sharing a common edge are defined to be at distance 1. The *distance to the boundary* of a face will be its distance to the exterior of the diagram considered as a face, i.e. a boundary face is at distance 1 from the boundary.

A diagram is said to be A-*narrow* if any 2-cell is at distance at most A from the boundary.

It is well-known, and we show in Appendix B in the form we need, that a linear isoperimetry implies narrowness of minimal diagrams.

2 The standard case: F_m

We now proceed to the proof of Gromov's now classical theorem (Theorem 1) that a random quotient of the free group F_m is trivial in density greater than $1/2$, and non-elementary hyperbolic in density smaller than this value.

We include this proof here because, first, it can serve as a useful template for understanding the general case, and, second, it seems that no completely correct proof has been published so far.

Recall that in this case, we consider a random quotient of the free group F_m on m generators by $(2m-1)^{d\ell}$ uniformly chosen *cyclically reduced* words of length ℓ.

A random cyclically reduced word is chosen in the following way: first choose the first letter ($2m$ possibilities), then choose the next letter in such a way that it is not equal to the preceding one ($2m-1$ possibilities), up to the last letter which has to be distinct both from the penultimate letter and the first one (which lets $2m-2$ or $2m-1$ choices depending on whether the penultimate letter is the same as the first one). The difference between $2m$ and $2m-1$ at the first position, and between $2m-1$ and $2m-2$ at the last position is negligible (as $\ell \to \infty$) and we will do as if we had $2m-1$ choices for each letter exactly.

So, for the sake of simplicity of the exposition, in the following we may assume that there are exactly $(2m-1)^\ell$ reduced words of length ℓ, with $2m-1$ choices for each letter. Bringing the argument to full correctness is a straightforward exercise.

2.1 Triviality for $d > 1/2$

The triviality of the quotient for $d > 1/2$ reduces essentially to the well-known

Probabilistic pigeon-hole principle – *Let $\varepsilon > 0$ and put $N^{1/2+\varepsilon}$ pigeons uniformly at random among N pigeon-holes. Then there are two pigeons in the same hole with probability tending to 1 as $N \to \infty$ (and this happens arbitrarily many times with growing N).*

Now, take as your pigeon-hole the word made of the first $\ell - 1$ letters of a random word of length ℓ. There are $(2m-1)^{\ell-1}$ pigeon-holes and we pick up $(2m-1)^{d\ell}$ random words with $d > 1/2$. Thus, with probability arbitrarily close to 1 with growing ℓ, we will pick two words of the form wa_i, wa_j where $|w| = \ell - 1$ and $a_i, a_j \in S$. Hence in the quotient group we will have $a_i = a_j$.

But as d is strictly bigger than $1/2$, this will not only occur once but arbitrarily many times as $\ell \to \infty$, with at each time a_i and a_j being chosen at random from S. That is, for big enough ℓ, all couples of generators $a, b \in S$ will satisfy $a = b$ in the quotient group. As S is symmetric, in particular they will satisfy $a = a^{-1}$.

The group presented by $\langle (a_i) \ ; \ a_i = a_i^{-1}, a_i = a_j \ \forall i, j \rangle$ is $\mathbb{Z}/2\mathbb{Z}$. In case ℓ is even this is exactly the group we get (as there are only relations of even length), and if ℓ is odd any relation of odd length turns $\mathbb{Z}/2\mathbb{Z}$ into $\{e\}$.

This proves the second part of Theorem 1.

2.2 Hyperbolicity for $d < 1/2$

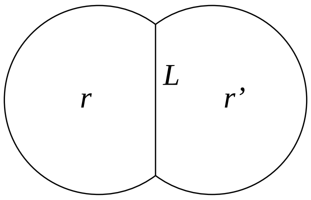
We proceed as follows: We will show that the only (reduced) daDd's which are fulfillable by a random presentation are those which satisfy some linear isoperimetric inequality. This is stronger than proving that only minimal diagrams satisfy an isoperimetric inequality: in fact, *all* reduced diagrams in a random group satisfy this inequality. (Of course this cannot be true of non-reduced diagrams since one can, for example, take any relator r and arrange an arbitrarily large diagram of alternating r's and r^{-1}'s like in a chessboard.)

Thus we will evaluate the probability that a given decorated abstract Dehn diagram can be fulfilled by a random presentation. We show that if the daDd violates the isoperimetric inequality, then this probability is very small and in fact decreases exponentially with ℓ.

Then, we apply the Cartan-Hadamard-Gromov-Papasoglu theorem for hyperbolic spaces, which tells us that to ensure hyperbolicity of a group, it is not necessary to check the isoperimetric inequality for *all* diagrams but for a *finite number* of them (see section A for details).

Say is it enough to check all diagrams with at most K faces, where K is some constant depending on d but not on ℓ. Assume we know that for each of these diagrams which violates the isoperimetric inequality, the probability that it is fulfillable decreases exponentially with ℓ. Let $D(K)$ be the (finite) number of daDd's with at most K faces, violating the isoperimetric inequality. The probability that at least one of them is fulfillable is less that $G(K)$ times some quantity decreasing exponentially with ℓ, and taking ℓ large enough ensures that with probability arbitrarily close to one, none of these daDd's is fulfillable. The conclusion then follows by the Cartan-Hadamard-Gromov-Papasoglu theorem.

The basic picture is as follows: Consider a daDd made of two faces of perimeter ℓ meeting along L edges. The probability that two given random relators r, r' fulfill this diagram is at most $(2m-1)^{-L}$, which is the probability that L given letters of r are the inverses of L given letters of r'. (Remember that as the relators are taken reduced, there are only $2m-1$ choices for each letter expect for the first one. As $2m-1 < 2m$ we can safely treat the first letter like the others, as doing otherwise would still sharpen our evaluation.)



Now, there are $(2m-1)^{d\ell}$ relators in the presentation. As we said, the probability that two given relators fulfill the diagram is at most $(2m-1)^{-L}$. Thus, the probability that there exist two relators in the presentation fulfilling the diagram is at most $(2m-1)^{2d\ell}(2m-1)^{-L}$, with the new factor accounting for the choice of the two relators.

This evaluation becomes non-trivial for $L > 2d\ell$. Observe that the boundary length of the diagram is $2\ell - 2L = 2(1-2d)\ell - 2(L - 2d\ell)$. That is, if $L \leqslant 2d\ell$ then the boundary is longer than $2(1-2d)\ell$, and if $L > 2d\ell$ then the probability that the diagram can be fulfilled is exponentially small with ℓ.

To go on with our intuitive reasoning, consider a graph with n relators instead of two. The number of "conditions" imposed by the graph is equal to the total length L of its internal edges, that is, the probability that a random assignment of relators satisfy them is $(2m-1)^{-L}$, whereas the number of choices for the relators is $(2m-1)^{nd\ell}$ by definition. So if $L > nd\ell$ the probability is too small. But if $L \leqslant nd\ell$, then the boundary length, which is equal to $n\ell - 2L$, is bigger than $(1-2d)n\ell$ which is the isoperimetric inequality we were looking for.

This is the picture we will elaborate on. In fact, what was false in the last paragraph is that if the same relator is to appear several times in the diagram, then we cannot simply multiply probabilities as we did, as these probabilities are no more independent.

Thus, let D be a reduced daDd. We will evaluate the probability that it can be fulfilled by relators of a random presentation.

Note that the original proof of Gromov forgot to deal with the case when two faces of the diagram bear the same relator. If all relators are distinct, all the probabilities are independent and the proof is easier. However, if two faces bear the same relator, then the probabilities that these faces stick to their neighbours are not independent, and we cannot simply multiply probabilities as in the basic picture.

Each face of D bears a number between 1 and $|D|$. Let n be the number of distinct numbers the faces bear in D. Of course, $n \leqslant |D|$. (This amounts to the case proved by Gromov if $n = |D|$.) Suppose, for simplicity, that these n distinct numbers are $1, 2, \ldots, n$.

To fulfill D is to give n relators $r_1, \ldots, r_n$ satisfying the relations imposed by the diagram.

We will construct an auxiliary graph Γ summarizing all letter relations imposed by the diagram D. Vertices of Γ will represent the letters of $r_1, \ldots, r_n$, and edges of Γ will represent inverseness (or equality, depending on orientation) of letters imposed by shared edges between faces of D.

Thus, take $n\ell$ vertices for Γ, arranged in n parts of ℓ vertices. Call the vertices corresponding to the faces of D bearing number i the i-th part of the graph. Each part is made of ℓ vertices.

We now explain what to take as edges of Γ.

In the diagram, every face is marked with a point on its boundary, and an orientation. Label the edges $1, 2, \ldots, \ell$ starting at the marked point, following the given orientation.

If, in the daDd D, the k-th edge of a face bearing number i is equal to the k'-th edge of an adjacent face bearing number j, then put an edge in Γ between the k-th vertex of the i-th part and the k'-th vertex of the j-th part. Decorate the newly added edge with -1 if the two faces' orientations agree, or with $+1$ if they disagree.

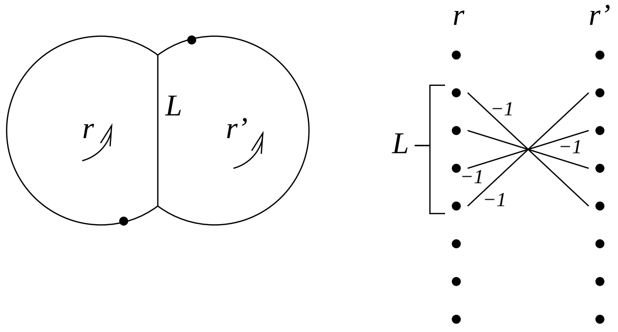
Thus, a -1 edge between the k-th vertex of the i-th part and the k'-th vertex of the j-th part means that the k-th letter of relator r_i has to be the inverse of the k'-th letter of relator r_j.

Successively add an edge to Γ in this way for each interior edge of the daDd D, so that the total number of edges of Γ is equal to the number of interior edges of D.

As D is reduced, the graph Γ can contain no loop. It may well have multiple edges, if, in the daDd, several pairs of adjacent faces bear the same numbers and have the same common edge.

Note that this graph only depends on the daDd D and in no way on the random presentation.

The graph Γ for the basic picture above is:



Now let us evaluate the probability that D is fulfillable. To fulfill D is to assign a generator to each vertex of Γ and see if the relations imposed by the edges are satisfied.

Remark that if the generator of any vertex of the graph is assigned, then this fixes the generators of its whole connected component. (And, maybe, depending on the signs of the edges of Γ, there is no correct assignation at all.) Thus, the number of degrees of freedom is at most equal to the number of connected components of Γ.

Thus (up to our approximation on the number of cyclically reduced words), the number of random assignments of cyclically reduced words to the vertices of Γ is $(2m-1)^{n\ell}$, whereas the number of those assignments satisfying the constraints of the edges is at most $(2m-1)^C$ where C is the number of connected components. Hence, the probability that a given assignment of n random words to the vertices of Γ satifies the edges relations is at most $(2m-1)^{C-n\ell}$.

This is the probability that n *given* relators of a random presentation fulfill the diagram. Now there are $(2m-1)^{d\ell}$ relators in a random presentation, so the probability that we can find n of them fulfilling the diagram is at most $(2m-1)^{nd\ell}\,(2m-1)^{C-n\ell}$.

Now let Γ_i be the subgraph of Γ made of those vertices corresponding to a face of D bearing a number $\leqslant i$. Thus $\Gamma_1 \subset \Gamma_2 \subset \ldots \subset \Gamma_n = \Gamma$. Of course, the probability that Γ is fulfillable is less than any of the probabilities that Γ_i is fulfillable for $i \leqslant n$.

The above argument on the number of connected components can be repeated for Γ_i: the probability that Γ_i is fulfillable is at most $(2m-1)^{id\ell+C_i-i\ell}$ where C_i is the number of connected components of Γ_i.

This leads to setting

$$d_i = id\ell + C_i - i\ell$$

and following Gromov we interpret this number as the dimension of Γ_i, or, better, the dimension of the set of random presentations for which there exist i relators satisfying the conditions imposed by Γ_i. Thus:

$$\Pr(D \text{ is fulfillable}) \leqslant (2m-1)^{d_i} \qquad \forall i$$

Now turning to isoperimetry. Let m_i be the number of faces of D bearing number i. A vertex in the i-th part of Γ is thus of multiplicity at most m_i. Let A be the number of edges in Γ. We have

$$|\partial D| = |D|\,\ell - 2A = \ell \sum m_i - 2A$$

Thus we want to show that either the number of edges is small, or the fulfillability probability is small. The latter grows with the number of connected components of Γ, so this looks reasonable.

Let A_i be the number of edges in Γ_i. We now show that

$$A_{i+1} - A_i + m_{i+1}(d_{i+1} - d_i) \leqslant m_{i+1} d\ell$$

or equivalently that

$$A_{i+1} - A_i + m_{i+1}(C_{i+1} - (C_i + \ell)) \leqslant 0$$

Depart from Γ_i and add the new vertices and edges of Γ_{i+1}. When adding the ℓ vertices, the number of connected components increases by ℓ. So we only have to show that when adding the edges, the number of connected components decreases at least by $1/m_{i+1}$ times the number of edges added.

Call *external point* a point of $\Gamma_{i+1} \setminus \Gamma_i$ which shares an edge with a point of Γ_i. Call *internal point* a point of $\Gamma_{i+1} \setminus \Gamma_i$ which is not external. Call *external edge* an edge between an external point and a point of Γ_i, *internal edge* an edge between two internal points, and *external-internal edge* an edge between an external and internal point. Call *true internal point* a point which has at least one internal edge.

While adding the external edges, each external point is connected to a connected component inside Γ_i, and thus the number of connected components decreases by 1 for each external point.

Now add the internal edges (but not yet the external-internal ones): If there are N true internal points, these make at most $N/2$ connected components after adding the internal edges, so the number of connected components has decreased by at least $N/2$.

After adding the external-internal edges the number of connected components still decreases. Thus it has decreased by at least the number of external points plus half the number of true internal points.

Now as each external point is of degree at most m_{i+1}, the number of external plus external-internal edges is at most m_{i+1} times the number of external points. If there are N true internal points, the number of internal edges is at most $Nm_{i+1}/2$ (each edge is counted 2 times). So the total number of edges is at most m_{i+1} times the number of external points plus half the number of true internal points, which had to be shown.

Thus we have proved that $A_{i+1}-A_i+m_{i+1}(d_{i+1}-d_i) \leqslant m_{i+1}d\ell$. Summing over i yields

$$A + \sum m_i(d_i - d_{i-1}) \leqslant d\ell \sum m_i$$

Thus,

$$\begin{aligned}
|\partial D| &= \ell \sum m_i - 2A \\
&\geqslant \ell \sum m_i - 2d\ell \sum m_i + 2 \sum m_i (d_i - d_{i-1}) \\
&= \ell\,|D|\,(1-2d) + 2\sum d_i (m_i - m_{i+1})
\end{aligned}$$

But we can choose the order of the construction, and we may suppose that the m_i's are non-increasing, i.e. that we began with the relator appearing the biggest number of times in D, etc., so that $m_i - m_{i+1}$ is non-negative.

If all d_i's are non-negative, then we have the isoperimetric inequality $|\partial D| \geqslant \ell\,|D|\,(1-2d)$.

If some d_i is negative, we use the fact established above that the probability that the diagram is fulfillable is less than $(2m-1)^{\inf d_i}$. As $\sum m_i = |D|$, we have $\sum d_i(m_i - m_{i+1}) \geqslant |D| \inf d_i$. Thus $|\partial D| \geqslant \ell\,|D|\,(1-2d+2\inf d_i/\ell)$.

If $\inf d_i \geqslant -\ell(1-2d)/4$, we get the inequality $|\partial D| \geqslant \ell\,|D|\,(1/2-d)$ (hence the interest of taking $d < 1/2$...)

Otherwise, if $\inf d_i < -\ell(1-2d)/4$, the probability that D is fulfillable is less than $(2m-1)^{-\ell(1/2-d)/2}$.

Thus we have shown that: if D is a daDd, then either D satisfies the isoperimetric inequality

$$|\partial D| \geqslant \ell\,|D|\,(1/2-d)$$

or

$$\Pr(D \text{ is fulfillable}) \leqslant (2m-1)^{-\ell(1/2-d)/2}$$

(Observe the latter probability decreases exponentially with ℓ.)

In order to show that the group is hyperbolic, we have to show that the probability that there exists a daDd violating the isoperimetric inequality tends to 0 when $\ell \to \infty$. But here we use the local-to-global principle for hyperbolic grometry (or Cartan-Hadamard-Gromov-Papasoglu theorem, see Appendix A), which can be stated as:

Proposition – *For each $\alpha > 0$, there exist an integer $K(\alpha) \geqslant 1$ and an $\alpha' > 0$ such that, if a group is given by relations of length ℓ for some ℓ and if any reduced Dehn diagram with at most K faces satifies*

$$|\partial D| \geqslant \alpha\ell\,|D|$$

then any reduced Dehn diagram D satisfies

$$|\partial D| \geqslant \alpha'\ell\,|D|$$

(hence the group is hyperbolic).

Now take $\alpha = 1/2 - d$ and the K given by the proposition. If $N(K,\ell)$ is the number of daDd's with at most K faces and each face has ℓ edges, then the probability that one of them is fulfillable and violates the isoperimetric inequality is at most $N(K,\ell)\,(2m-1)^{-\ell(1/2-d)/2}$.

Let us evaluate $N(K,\ell)$. As the relators in the presentation are taken to be cyclically reduced, we only have to consider regular diagrams (see 1). A regular daDd is only a planar graph with some decoration on the edges, namely, a planar graph with on each edge a length indicating the number of edges of the daDd it represents, and with vertices of degree at least 3 (and, as in a daDd, every face is decorated with a starting point, an orientation, and a number between 1 and K). Let $G(K)$ be the number of planar graphs with vertex degree at least 3. In such a graph there are (by Euler's formula) at most $3K$ edges, so there are at most ℓ^{3K} choices of edge lengths, and we have $(2\ell K)^K$ choices for the decoration of each face (orientation, starting point and number between 1 and K).

So $N(K,\ell) \leqslant G(K)(2K)^K\ell^{4K}$. As this is polynomial in ℓ, the probability $N(K,\ell)\,(2m-1)^{-\ell(1/2-d)/2}$ tends to 0 as $\ell\to\infty$.

This proves that the quotient is hyperbolic; we now show that it is infinite. In fact we will prove the stronger property that it is torsion-free.

Indeed, the isoperimetric inequality above is not only valid for minimal diagrams, but for any reduced diagram.

Now suppose there is a non-trivial torsion element x of order $n>0$. This defines a reduced sphere in the Cayley complex of the group. A sphere violates any isoperimetric inequality as it has no border. Now the sphere is not exactly a Dehn diagram (it is not planar), but the above argument carries readily to those kind of limit Dehn diagrams of zero boundary length. Hence a contradiction.

This proves Theorem 1.

3 Outline of the argument

Here we explain some of the ideas of the proof of Theorems 2, 3 and 4.

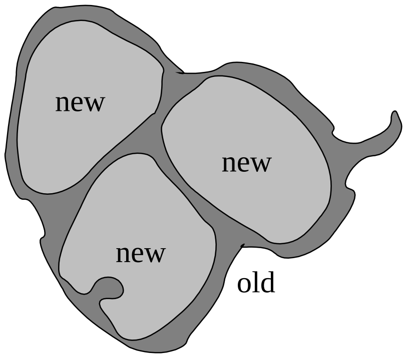
We will give a general theorem for hyperbolicity of random quotients by words taken from some probability measures on the set of all words. We will need somewhat technical axioms on the measures (for examples, that they weight only long words). Here we give a heuristic justification of why these axioms are needed.

We proceed by showing that Dehn diagrams of the quotient $G/\langle R\rangle$ satisfy a linear isoperimetric inequality.

If D is a Dehn diagram of the quotient, let D' be the subcomplex of D made of relators of the presentation of G ("old relators") and D'' the subcomplex made of relators in R ("new relators").

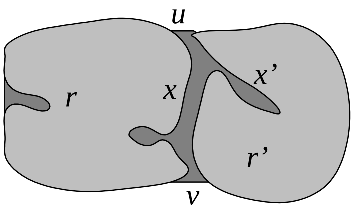
Say the new relators have length of order ℓ where ℓ is much bigger than the hyperbolicity constant of G. (This will be Axiom 1.)

The main point will be that D' is a diagram in the hyperbolic group G, and, as such, is narrow (see Appendix B). We show below that its narrowness is of order $\log \ell$. Hence, if ℓ is big enough, the diagram D can be viewed as big faces representing the new relators, separated by a thin layer of "glue" representing the old relators. The "glue" itself may contain invaginations in the new relators and narrow excrescences on the boundary.



3.1 A basic picture

As an example, let us study a basic picture consisting of two new relators separated by some old stuff. Say that two random new relators r, r' are "glued" along subwords of length L, L' (we may have $L \neq L'$). Let w be the word bordering the part of the diagram made of old relators, we have $|w| = L + L' + o(\ell)$. By construction, w is a word representing the trivial element in G. Write $w = xux'v$ where x is a subword of r of length L, x' is a subword of r' of length L', and u and v are short words.



Let us evaluate the probability that such a diagram exists. Take two given random relators r, r' in R. The probability that they can be glued along subwords x, x' of lengths L, L' by narrow glue in G is the probability that there exist short words u, v such that $xux'v = e$ in G.

If (as in the standard case) there were no glue (no old relators) and r and r' were uniformly chosen random reduced words, the probability that r and r' could be glued along subword x, x' of length L (we would have $L = L'$ in this case) would be $(2m)^{-L}$. But we now have to consider the case when then x and x' are equal, not as words, but as elements of G (and up to small words u and v, which we will neglect).

If, for example, the relators are uniformly chosen random words, then x and x' are independent subwords, and the probability that x and x' are (almost) equal in G is the probability that $x{x'}^{-1} = e$; but $x{x'}^{-1}$ is a uniformly chosen random word, and by definition the probability that it is equal to e is controlled by the gross cogrowth of G: this is roughly $(2m)^{-(1-\theta)(L+L')}$ (recall the alternate definition of gross cogrowth in section 1.2).

In order to deal not only with uniformly chosen random words but with other situations such as random geodesic words, we will need a control on the probability that two relators can be glued (modulo G) along subwords of length L and L'. This will be our Axiom 3: we will ask this probability to decrease like $(2m)^{-\beta(L+L')}$ for some exponent β (equal to $1-\theta$ for plain random words).

Now in the simple situation with two relators depicted above, the length of the boundary of the diagram is not exactly $2\ell - L - L'$, since there can be invaginations of the relators, i.e. long part of the relators which are equal to short elements in G (as in the left part of the picture above). In the case of uniformly chosen random relators, by definition the probability that a part of length L of a relator is (nearly) equal to e in G is roughly $(2m)^{-(1-\theta)L}$. So, again inspired by this case, we will ask for an axiom controlling the length of subwords of our relators. This will be our Axiom 2.

Axiom 4 will deal with the special case when $r = r'^{-1}$, so that the words x and x' above are equal, and not at all chosen independently as we implicitly assumed above. In this case, the virtual centre of the group will matter.

This was for given r and r'. But there are $(2m)^{d\ell}$ relators in R, so we have $(2m)^{2d\ell}$ choices for r, r'. Thus, the probability that in R, there are two new relators that glue along subwords of length L, L' is less than $(2m)^{2d\ell}(2m)^{-\beta(L+L')}$.

Now, just observe that the length of the boundary of the diagram is (up to the small words u and v) $2\ell - L - L'$. On the other hand, when $d < \beta$, the exponent $2d\ell - \beta(L+L')$ of the above probability will be negative as soon as $L + L'$ is bigger than 2ℓ. This is exactly what we want to prove: either the boundary is big, or the probability of existence of the diagram is small.

This is comparable to the former situation with random quotients of the free group: in the free group, imposing two random relators to glue along subwords of lengths L and $L' = L$ results in L "equations" on the letters. For example, in the case of plain random words, in a group of gross cogrowth θ, imposing two random words to glue along subwords of lengths L, L' results in $\beta(L+L')$ "equations" on these random words with $\beta = 1 - \theta$.

Now for diagrams having more than two new relators, essentially the number of "equations" imposed by the gluings is β times the total internal length of the relators. The boundary is the external length. If there are n new relators and the total internal length is A, then the boundary is roughly $n\ell - A$. But the probability of existence of such a diagram is $(2m)^{-\beta A}(2m)^{nd\ell}$ where the last factor accounts for the choice of the n relators among the $(2m)^{d\ell}$ relators of R. So if $d < \beta$, as soon as $A > n\ell$, the probability decreases exponentially with ℓ.

3.2 Foretaste of the Axioms

As suggested by the above basic picture, we will demand four axioms: one saying that our random relators are of length roughly ℓ, another saying that subwords of our relators are not too short, another one controlling the probability that two relators glue along long subwords (that is, the probability that these subwords are nearly equal in G), and a last one controlling the probability that a relator glues along its own inverse.

As all our estimates are asymptotic in the length of the words considered, we will be allowed to apply them only to sufficiently long subwords of our relators (and not to one individual letter, for example), that is, to words of length at least $\varepsilon\ell$ for some ε.

Note that in order to be allowed to apply these axioms to any subword of the relators at play, whatever happens elsewhere, we will need to ask that different subwords of our relators behave quite independently from each other; in our axioms this will result in demanding that the probability estimates hold for a subword of a relator conditionnally to whatever the rest of the relator is.

This is a strong independence condition, but, surprisingly enough, is it valid not only for uniformly chosen random words (where by definition everything is independent, in any group), but also for randomly chosen geodesic words. This is a specific property of hyperbolic groups.

Several exponents will appear in the axioms. As we saw in the basic picture, the maximal density up to which the quotient is non-trivial is exactly the minimum of these exponents. Back to the intuition behind the density model of a random quotient (see the introduction), the exponents in our axioms indicate how many equations it takes in G to have certain gluings in our relators, whereas the density of the random quotient is a measure of how many equations we can reasonably impose so that it is still possible to find a relator satisfying them among our randomly chosen relators. So this intuition gets a very precise numerical meaning.

4 Axioms on random words implying hyperbolicity of a random quotient, and statement of the main theorem

We want to study random quotients of a (non-elementary) hyperbolic group G by randomly chosen elements. Let μ_ℓ be the law, indexed by some parameter ℓ to tend to infinity, of the random elements considered.

We will always assume that μ_ℓ is a symmetric measure, i.e. for any $x \in G$, we have $\mu_\ell(x) = \mu_\ell(x^{-1})$.

We will show that if these measures satisfy some simple axioms, then the random quotient by elements picked under these measures is hyperbolic.

For each of the elements of G weighted bu μ_ℓ, fix once and for all a representation of it as a word (and choose inverse words for inverse elements), so that μ_ℓ can be considered as a measure on words. Satisfaction of our axioms may depend on such a choice.

Let μ_ℓ^L be the law μ_ℓ restricted (and rescaled) to words of length L (or 0 if there are no such words in the support of μ). In most applications, μ_ℓ will weight only words of length ℓ, but we will occasionally use laws μ_ℓ weighting words of length comprised between, say, $A\ell$ and $B\ell$.

To pick a random set R of density at most d is to pick, for each length L, independently, at most $(2m)^{dL}$ random words of length L according to law μ_ℓ^L. That is, for each length, the density is at most d.

(We say "at most" because we do not require that exactly $(2m)^{dL}$ words of length L are taken for each L. Taking smaller R will result in a hyperbolic quotient as well.)

We want to show that if d is less than some quantity depending on μ_ℓ (and G, since μ_ℓ takes value in G), then the random quotient $G/\langle R\rangle$ is very probably non-elementary hyperbolic.

4.1 Asymptotic notations

By the notation $f(\ell) \approx g(\ell)$ we shall mean that

$$\lim_{\ell\to\infty} \frac{1}{\ell} \log f(\ell) = \lim_{\ell\to\infty} \frac{1}{\ell} \log g(\ell)$$

We define the notation $f(\ell) \lesssim g(\ell)$ similarly. We will say, respectively, that f is roughly equal or roughly less than g.

4.2 Some vocabulary

Here we give technical definitions designed in such a manner that the axioms can be stated in a natural way. We recommend to look at the axioms first.

Let x be a word. For each a, b in $[0;1]$ such that $a+b \leqslant 1$, we denote by $x_{a;b}$ the subword of x going from the $(a\,|x|)$-th letter (taking integer part, and inclusively) to the $((a+b)\,|x|)$-th letter (taking integer part, and exclusively), so that a indicates the position of the subword, and b its length. If $a+b>1$ we cycle around x.

Definition 6 – *Let P be a property of words. We say that*

$$\Pr(P) \lesssim p(\ell)$$

for any subword under μ_ℓ if for any $a, b \in [0;1]$, whenever we pick a word x according to μ_ℓ we have

$$\Pr\left(P(x_{a;b}) \mid x_{0;a}\right) \lesssim p(\ell) \quad \textit{if } a+b \leqslant 1$$

or

$$\Pr\left(P(x_{a;b}) \mid x_{a+b-1;a}\right) \lesssim p(\ell) \quad \textit{if } a+b>1$$

and if moreover the constants implied in $\lesssim$ *are uniform in* a*, and, for each* $\varepsilon > 0$*, uniform when* b *ranges in the interval* $[\varepsilon; 1]$.

That is, we pick a subword of a given length and ask the probability to be bounded independently of whatever happened in the word up to this subword (if the subword cycles around the end of the word, we condition by everything not in the subword).

It would not be reasonable to ask that the constants be independent of b for arbitrarily small b. For example, if μ_ℓ consists in choosing uniformly a word of length ℓ, then taking $b = 1/\ell$ amounts to considering subwords of length 1, which we are unable to say anything interesting about.

We give a similar definition for properties depending on two words, but we have to beware the case when they are subwords of the same word.

Definition 7 – *Let* P *be a property depending on two words. We say that*

$$\Pr(P) \lesssim p(\ell)$$

for any two disjoint subwords under μ_ℓ *if for any* $a, b, a', b' \in [0; 1]$ *such that* $a+b \leqslant 1$, $a'+b' \leqslant 1$*, whenever we pick two independent words* x, x' *according to* μ_ℓ *we have*

$$\Pr\left(P(x_{a;b},\, x'_{a';b'}) \mid x_{0;a},\, x'_{0;a'}\right) \lesssim p(\ell)$$

and if for any $a, b, a', b' \in [0; 1]$ *such that* $a \leqslant a+b \leqslant a' \leqslant a'+b' \leqslant 1$*, whenever we pick a word* x *according to* μ_ℓ*, we have*

$$\Pr\left(P(x_{a,b},\, x_{a';b'}) \mid x_{0;a},\, x_{a+b;a'}\right) \lesssim p(\ell)$$

We give similar definitions when $a+b>1$ *or* $a'+b'>1$*, conditioning by every subword not in* x *or* x'.

Furthermore, we demand that the constants implied in $\lesssim$ *be uniform in* a, a'*, and, for each* $\varepsilon > 0$*, uniform when* b, b' *range in the interval* $[\varepsilon; 1]$.

We are now ready to express the axioms we need on our random words.

4.3 The Axioms

Our first axiom states that μ_ℓ consists of words of length roughly ℓ up to some constant factor. This is crucial for the hyperbolic local-to-global principle (Appendix A).

Axiom 1 – *There is a constant* κ_1 *such that* μ_ℓ *weights only words of length between* ℓ/κ_1 *and* $\kappa_1\ell$.

Note this axiom applies to words picked under μ_ℓ, and not especially subwords, so it does no rely on our definitions above. But of course, if $|x| \leqslant \kappa_1 \ell$, then $|x_{a;b}| \leqslant b\kappa_1 \ell$.

Our second axiom states that subwords do not probably represent short elements of the group.

Axiom 2 – *There are constants κ_2, β_2 such that for any subword x under μ_ℓ, for any $\varepsilon \leqslant 1$, we have*

$$\Pr\left(\|x\| \leqslant \kappa_2\,|x|\,(1-\varepsilon)\right) \lesssim (2m)^{-\beta_2\varepsilon|x|}$$

uniformly in ε.

Our next axiom controls the probability that two subwords are almost inverse in the group. We will generally apply it with $n(\ell) = O(\log \ell)$.

Axiom 3 – *There are constants β_3 and γ such that for any two disjoint subwords x, y under μ_ℓ, the probability that there exist words u and v of length at most $n = n(\ell)$, such that $xuyv = e$ in G, is roughly less than $(2m)^{\gamma n}(2m)^{-\beta_3(|x|+|y|)}$.*

Our last axiom deals with algebraic properties of commutation with short words.

Axiom 4 – *There exist constants β_4 and γ such that, for any subword x under μ_ℓ, the probability that there exist words u and v of length at most $n = n(\ell)$, such that $ux = xv$ and $u \neq e$, $v \neq e$, is roughly less than $(2m)^{\gamma n}(2m)^{-\beta_4|x|}$*

If G has a non-trivial virtual centre, this axiom will probably fail to be true, as if u lies in the virtual centre, then $x^{-1}ux$ will be short.

We will see below (section 4.5) that, in a hyperbolic group without virtual centre, the algebraic Axiom 4 is a consequence of Axioms 1 and 3 combined with a more geometric axiom which we state now.

Axiom 4' – *There are constants β_4' and γ such that, for any $C > 0$, for any subword x under μ_ℓ, the probability that there exists a word u of length at most $n = n(\ell)$ such that some cyclic permutation x' of xu satisfies $\|x'\| \leqslant C \log \ell$, is roughly less than $(2m)^{\gamma n}(2m)^{-\beta_4'|x|}$.*

Remark 8 – Let μ_ℓ' be a family of measures such that $\mu_\ell' \lesssim \mu_\ell$. As our axioms consist only in rough upper bounds, if the family μ_ℓ satisfy them, then so does the family μ_ℓ'.

Note that as we condition every subword by whatever happened before, our axioms imply that subwords at different places are essentially independent. This is of course true of plain random words, but also of geodesic words and reduced words as we will see below.

4.4 The Theorem

Our main tool is the following

Theorem 9 – *Let G be a non-elementary hyperbolic group with trivial virtual centre. Let μ_ℓ be a family of symmetric measures indexed by ℓ, satisfying Axioms 1, 2, 3 and 4 (or 4'). Let R be a set of random words of density at most d picked under μ_ℓ.*

If $d < \min(\beta_2, \beta_3, \beta_4)$, then with probability exponentially close to 1 as $\ell \to \infty$, the random quotient $G/\langle R\rangle$ is non-elementary hyperbolic, as well as all the intermediate quotients $G/\langle R'\rangle$ with $R' \subset R$.

Section 6 is devoted to the proof.

Remark 10 – Remark 8 tells that if the theorem applies to a family of measures μ_ℓ, it applies as well to all families of measures $\mu'_\ell \lesssim \mu_\ell$.

4.5 On Axiom 4

We show here that in a hyperbolic group without virtual centre, Axioms 1, 3 and 4' imply Axiom 4. The proof makes the algebraic nature of this axiom clear: in a hyperbolic group, it means that subwords under μ_ℓ are probably not torsion elements, neither close to powers of short elements..

Let μ_ℓ be a measure satisfying Axioms 1, 3 and 4'. Let $\beta = \min(\beta_3, \beta'_4)$.

Proposition 11 – *The probability that, for a subword x under μ_ℓ, there exists a word u of length at most $n = n(\ell)$ such that xu is a torsion element, is roughly less than $(2m)^{\gamma n}(2m)^{-\beta|x|}$.*

Proof – In a hyperbolic group, there are only finitely many conjugacy classes of torsion elements (see [GH], p. 73). Let L be the maximal length of a shortest element of a conjugacy class of torsion elements, we have $L < \infty$. Now every torsion element is conjugated to an element of length at most L.

Suppose xu is a torsion element. It follows from Corollary 49 that some cyclic permutation of it is conjugate to an element of length at most L by some word of length at most $\delta \log_2 |xu| + C'_c + 1$ where C'_c is a constant depending on the group. In particular, this cyclic conjugate has norm at most $L + 2(\delta \log_2 |xu| + C'_c + 1)$.

Suppose, by Axiom 1, that $|x| \leqslant \kappa_1 \ell$.

There are $|xu| \leqslant \kappa_1 \ell + n$ cyclic conjugates of xu. The choice of the cyclic conjugate therefore only introduces a polynomial factor in ℓ. Let x' denote the cyclic conjugate of xu at play.

Thus we have to evaluate the probability that $\|x'\| \leqslant L + 2(\delta \log_2 |x'| + C'_c + 1)$. As L and C'_c are mere constants, Axiom 4' precisely says that this probability is roughly less than $(2m)^{\gamma n}(2m)^{-\beta'_4|x|}$. □

Proposition 12 – *Let $w \in G$. For any subword x under μ_ℓ, the probability that $x = w$ in G is roughly less than $(2m)^{-\beta_3|x|}$ (uniformly in w).*

Proof – Suppose that the probability that a subword x under μ_ℓ is equal to w is equal to p. Then, by symmetry, the probability that an independent disjoint subword y with $|y| = |x|$ is equal to w^{-1} is equal to p as well. So the probability that two disjoint subwords x and y are inverse is at least p^2. But Axiom 3 tells (taking $u = v = e$) that this probability is roughly at most $(2m)^{-\beta_3(|x|+|y|)} = (2m)^{-2\beta_3|x|}$, hence $p \lesssim (2m)^{-\beta_3|x|}$. □

Recall that the virtual centre of a hyperbolic group is the set of elements whose action on the boundary at infinity is trivial. Recall we set $\beta = \min(\beta_3, \beta_4')$. We now prove the satisfaction of Axiom 4.

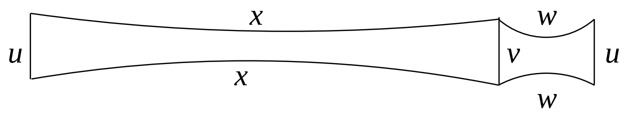
Proposition 13 – *Suppose G has trivial virtual centre. There is a constant γ such that for any subword x under μ_ℓ, the probability that there exist words u, v of length at most $n = n(\ell)$, such that $ux = xv$ in G, with u, v not equal to e, is roughly less than $(2m)^{\gamma n - \beta|x|}$.*

So Axiom 4 is satisfied with $\beta_4 = \min(\beta_3, \beta_4')$.

Proof – Denote by x again a geodesic word equal to x in G.

The words u and v are conjugate (by x), and are of length at most n. After Corollary 49 they are conjugate by a word w of length at most Cn where C is a constant depending only on G.

Let us draw the hyperbolic quadrilateral $xwuw^{-1}x^{-1}u^{-1}$. This is a commutation diagram between xw and u.



The word xw may or may not be a torsion element. The probability that there exists a word w of length at most Cn, such that xw is a torsion element, is roughly less than $(2m)^{\gamma Cn - \beta|x|}$ by Proposition 11. In this case we conclude.

Now suppose that xw is not a torsion element. Then we can glue the above diagram to copes of itself along their u-sides. This way we get two quasi-geodesics labelled by $((xw)^n)_{n\in\mathbb{Z}}$ that stay at finite distance from each other. The element u acting on the first quasi-geodesic gives the second one.

These two quasi-geodesics define an element $\tilde{x}$ in the boundary of G. This element is of course stabilized by xw, but it is stabilized by u as well. This means that either u is in the virtual centre (which we assumed was not the case), or that u is a hyperbolic element.

In the latter case, the idea is that xw and u are two elements of the stabilizer of the same point of the boundary of G. This means that xw is close to some power of the short element u, which is unlikely.

Let us use the same trick as above with the roles of xw and u exchanged: glue the diagram above to copies of itself by the (xw)-side. This defines two quasi-geodesics labelled by $(u^n)_{n\in\mathbb{Z}}$, one of which goes to the other when acted upon by xw.

Namely, let Δ be a geodesic equivalent to (u^n). As xw stabilizes the limit of Δ, the element xw sends Δ to Δ' where Δ' is equivalent to Δ. But two equivalent geodesics in a hyperbolic group stay at Hausdorff distance at most R where R is a constant depending only on the group (see [GH], p. 119).

Especially, the element xw is at distance at most R of the geodesic Δ (say that we translate everything in the Cayley graph so that Δ passes through the origin e, thus xw lies on Δ'). Let y be a point on Δ realizing this distance. As $|xw| \leqslant |x| + |w|$, we have $|y| \leqslant |x| + |w| + R$. There are at most $2|x| + 2|w| + 2R + 1$ such possible points on Δ (since Δ is a geodesic). For each of these points, the probability that x falls within distance $R + |w|$ of it is roughly less than $(2m)^{R+|w|}(2m)^{-\beta|x|}$ by Propositions 12 applied to all of these points. So the probability that x falls within distance less than $R + |w|$ of any one of the y's on a given geodesic Δ is roughly less than $(2|x| + 2|v| + 2R + 1)(2m)^{R+|w|}(2m)^{-\beta|x|}$ which in turn is roughly less than $(2m)^{Cn-\beta|x|}$ as $|w| \leqslant Cn$ and R is a constant.

But each different u defines a different Δ. There are at most $(2m)^{|u|} \leqslant (2m)^n$ possibilities for u. Finally, the probability that x falls within distance $R + |w|$ of any one of the geodesics defined by these u's is less than $(2m)^{n+Cn-\beta|x|}$ as was to be shown. $\square$

5 Applications of the main theorem

We now show how Theorem 9 leads, with some more work, to the theorems on random quotients by plain words, reduced words and geodesic words given in the introduction.

We have three things to prove:

- first, that these three models of a random quotient satisfy our axioms with the right critical densities;
- second, as Theorem 9 only applies to hyperbolic groups with trivial virtual centre, we have to find a way to get rid of the virtual centre;
- third, we have to prove triviality for densities above the critical one.

Once this is done, Theorems 2, 3 and 4 will be proved.

We will have to work differently if we consider quotients by plain random words, by random reduced words or by random geodesic words.

For instance, satisfaction of the axioms is very different for plain words and for geodesic words, because in plain random words, two given subwords fo the same word are chosen independently, which is not the case at all *a priori* for a geodesic word.

Furthermore, proving triviality of a quotient involves small scale phenomena, which are very different in our three models of random words (think of a random quotient of $\mathbb{Z}$ by random words of ℓ letters ± 1 or by elements of size exactly ℓ).

These are the reasons why the next three sections are divided in cases, and why we did not include these properties in a general and technical theorem such as Theorem 9.

Note that it is natural to express the critical densities in terms of the ℓ-th root of the total number of words of the kind considered, that is, in base $2m$ for plain words, $2m-1$ for reduced words and $(2m)^g$ for geodesic words.

5.1 Satisfaction of the axioms

5.1.1 The case of plain random words

We now take as our measure for random words the uniform measure on all words of length ℓ. Axiom 1 is satisfied by definition.

In this section, we denote by B_ℓ (as "Brownian") a random word of length ℓ uniformly chosen among all $(2m)^\ell$ possible words.

Recall θ is the gross cogrowth of the group, that is, the number of words of length ℓ which are equal to e in the group is roughly $(2m)^{\theta\ell}$ for even ℓ.

Recall the alternate definition of gross cogrowth given in the introduction: the exponent of return to e of the random walk in G is $1-\theta$. This is at the heart of what follows.

We will show that

Proposition 14 – *Axioms 1, 2, 3, 4' are satisfied by plain random uniformly chosen words, with exponent $1-\theta$ (in base $2m$).*

By definition, disjoint subwords of a uniformly taken random word are independent. So we do not have to care at all with the conditional probabilities of the axioms (contrary to the case of geodesic words below). Conditionnally to anything else, every subword x follows the law of $B_{|x|}$.

The definition of gross cogrowth only applies to even lengths. If ℓ is odd, either there are some relations of odd length in the presentation of the group, and then the limits holds, or there are no such relations, and the number of words of length ℓ equal to e is zero. In any case, this number is $\lesssim (2m)^{\theta\ell}$.

This is a delicate (but irrelevant) technical point: We should care with parity of the length of words. If there are some relations of odd length in our group, then the limit in the definition of gross cogrowth is valid regardless of parity of ℓ, but in general this is not the case (as is examplified by the free group). In order to get valid results for any length, we therefore often have to replace a $\approx$ sign with a $\lesssim$ one. In many cases, our statements of the form "$\Pr(\dots) \lesssim f(\ell)$" could in fact be replaced by "$\Pr(\dots) \approx f(\ell)$ if ℓ is even or if there are relations of odd length, and $\Pr(\dots) = 0$ otherwise". Here is the first example of such a situation.

Proposition 15 – *The probability that B_ℓ is equal to e is roughly less than $(2m)^{-(1-\theta)\ell}$.*

Proof – Alternate definition. □

Proposition 16 –

$$\Pr(\|B_\ell\| \leqslant \ell') \lesssim (2m)^{-(1-\theta)\left(\ell-\frac{\theta}{1-\theta}\ell'\right)}$$

uniformly in $\ell' \leqslant \ell$.

In particular, the escaping speed is at least $\frac{1-\theta}{\theta}$. So Axiom 2 is satisfied with $\kappa_2 = \frac{1-\theta}{\theta}$ and $\beta_2 = 1-\theta$.

Proof – For any L between 0 and ℓ', we have that $\Pr(B_{\ell+L} = e) \geqslant (2m)^{-L}\Pr(\|B_\ell\| = L)$.

But $\Pr(B_{\ell+L} = e) \lesssim (2m)^{-(1-\theta)(\ell+L)}$ (and this is uniform in $L \leqslant \ell$ since in any case, $\ell + L$ is at least equal to ℓ), hence the evaluation for a given L.

Now, summing over L between 0 and ℓ' introduces only a subexponential factor in ℓ. □

Proposition 17 – *The probability that, for two independently chosen words B_ℓ and $B'_{\ell'}$, there exist words u and v of length at most $n = n(\ell)$, such that $B_\ell u B'_{\ell'} v = e$ in G, is roughly less than $(2m)^{(2+2\theta)n}(2m)^{-(1-\theta)(\ell+\ell')}$.*

That is, Axiom 3 is satisfied with exponent $1-\theta$.

Proof – For any word u, we have $\Pr(B_{|u|} = u) \geqslant (2m)^{-|u|}$.

So let u and v be any two fixed words of length at most n. We have $\Pr(B_{\ell+|u|+\ell'+|v|} = e) \geqslant (2m)^{-|u|-|v|}\Pr(B_\ell u B'_{\ell'} v = e)$.

We know that $\Pr(B_{\ell+|u|+\ell'+|v|} = e) \lesssim (2m)^{-(1-\theta)(\ell+|u|+\ell'+|v|)}$.

So $\Pr(B_\ell u B'_{\ell'} v = e) \lesssim (2m)^{\theta(|u|+|v|)}(2m)^{-(1-\theta)(\ell+\ell')}$.

Now there are $(2m)^{|u|+|v|}$ choices for u and v. □

Proposition 18 – *The probability that there exists a word u of length at most $n = n(\ell)$, such that some cyclic conjugate of $B_\ell.u$ is of norm less than $C\log\ell$, is roughly less than $(2m)^{(1+\theta)n}(2m)^{-(1-\theta)\ell}$.*

So Axiom 4' is satisfied with exponent $1-\theta$.

Proof – As above, for any word u, we have $\Pr(B_{|u|} = u) \geqslant (2m)^{-|u|}$. So any property of $B_\ell.u$ occurring with some probability will occur for $B_{\ell+|u|}$ with at least $(2m)^{-|u|}$ times this probability. We now work with $B_{\ell+|u|}$.

Any cyclic conjugate of a uniformly chosen random word is itself a uniformly chosen random word, so we can assume that the cyclic conjugate at play is $B_{\ell+|u|}$ itself. There are $\ell + |u|$ cyclic conjugates, so the choice of the cyclic conjugate only introduces a subexponential factor in ℓ and $|u|$.

But we just saw above in Proposition 16 (when proving satisfaction of Axiom 2) that the probability that $\left\|B_{\ell+|u|}\right\| \leqslant L$ is roughly less than $(2m)^{-(1-\theta)\left(|u|+\ell-\frac{\theta}{1-\theta}L\right)}$.

Summing over the $(2m)^{|u|}$ choices for u yields the desired result, taking $L = C\log\ell$. □

So plain random words satisfy our axioms.

5.1.2 The case of random geodesic words

The case of geodesic words is a little bit more clever, as subwords of a geodesic word are not *a priori* independent.

For each element $x \in G$ such that $\|x\| = \ell$, fix once and for all a representation of x by a word of length ℓ. We are going to prove that when μ_ℓ is the uniform law on the sphere of radius ℓ in G, Axioms 1-4' are satisfied.

Recall that g is the growth of the group: by definition, the number of elements of length ℓ in G is roughly $(2m)^{g\ell}$. As G is non-elementary we have $g > 0$ (otherwise there is nothing to prove).

Proposition 19 – *Axioms 1, 2, 3, 4' are satisfied by random uniformly chosen elements of norm ℓ, with exponent $1/2$ (in base $(2m)^g$).*

Our proofs also work if μ_ℓ is the uniform measure on the spheres of radius between $\ell - L$ and $\ell + L$ for any fixed L. We will use this property later.

Note that Axioms 1 and 2 are trivially satisfied for geodesic words, with $\kappa_1 = \kappa_2 = 1$ and $\beta_2 = \infty$.

The main obstacle is that two given subwords of a geodesic word are not independent. We are going to replace the model of randomly chosen elements of length ℓ by another model with more independence, and prove that these two models are roughly equivalent.

Let X_ℓ denote a random uniformly chosen element on the sphere of radius ℓ in G. For any x on this sphere, we have $\Pr(X_\ell = x) \approx (2m)^{-g\ell}$.

Note that for any $\varepsilon > 0$, for any $\varepsilon\ell \leqslant L \leqslant \ell$ the rough evaluation of the number of points of length L by $(2m)^{gL}$ can by taken uniform for L in this interval (take ℓ so that $\varepsilon\ell$ is big enough).

First, we will change a little bit the model of random geodesic word. The axioms above use a strong independence property of subwords of the words taken. This independence is not immediately satisfied for subwords of a given random geodesic word (for example, in the hyperbolic group $F_2 \times \mathbb{Z}/2\mathbb{Z}$, the occurrence of a generator of order 2 somewhere prevents it from occurring anywhere else in a geodesic word). So we will cheat and consider an alternate model of random geodesic words.

For a given integer N, let X_ℓ^N be the product of N random uniformly chosen geodesic words of length ℓ/N. We will compare the law of X_ℓ to the law of X_ℓ^N.

Let $x \in G$ such that $\|x\| = \ell$. We have $\Pr(X_\ell = x) \approx (2m)^{-g\ell}$. Let $x = x_1 x_2 \dots x_N$ where each x_i is of length ℓ/N. The probability that the i-th segment of X_ℓ^N is equal to x_i is roughly $(2m)^{-g\ell/N}$. Multiplying, we get $\Pr(X_\ell^N = x) \approx (2m)^{-g\ell}$.

Thus, if P is a property of words, we have for any given N that

$$\Pr(P(X_\ell)) \lesssim \Pr(P(X_\ell^N))$$

(The converse inequality is false as the range of values of X_ℓ^N is not contained in that of X_ℓ.)

Of course, the constants implied in $\lesssim$ depend on N. We are stating that for any fixed N, when ℓ tends to infinity the law of the product of N words of length ℓ/N encompasses the law of X_ℓ, and *not* that for a given ℓ, when N tends to infinity the law of N words of length ℓ is close to the law of a word of length $N\ell$, which is false.

We are going to prove the axioms for X_ℓ^N instead of X_ℓ. As the axioms all state that the probability of some property is roughly less than something, these evaluations will be valid for X_ℓ.

The N to use will depend on the length of the subword at play in the axioms. With notations as above, if $x_{a;b}$ is a subword of length $b\ell$ of X_ℓ, we will choose an N such that ℓ/N is small compared to $b\ell$, so that $x_{a;b}$ can be considered the product of a large number of independently randomly chosen smaller geodesic words. This is fine as our axioms precisely *do not* require the evaluations to be uniform when the relative length b tends to 0.

First, we need to study multiplication by a random geodesic word.

Let $(x|y)$ denote the Gromov product of the two elements $x, y \in G$. That is, $(x|y) = \frac{1}{2}\left(\|x\| + \|y\| - \|x^{-1}y\|\right)$.

Proposition 20 – *Let $x \in G$ and $L \leqslant \ell$. We have*

$$\Pr\left((x|X_\ell) \geqslant L\right) \lesssim (2m)^{-gL}$$

uniformly in $L \leqslant \ell$.

Proof – Suppose $(x|X_\ell) = L'$ for some $L \leqslant L' \leqslant \ell$. Let y be the point at distance L' on a geodesic joining e to x. By the triangle-tripod transformation in exX_ℓ, the equality $(x|X_\ell) = L'$ means that X_ℓ is at distance at most $\ell - L' + 4\delta$ from y. There are roughly at most $(2m)^{g(\ell-L'+4\delta)}$ such points. Thus, the probability that X_ℓ is equal to one of them is roughly less than $(2m)^{g(\ell-L'+4\delta)-g\ell} \approx (2m)^{-gL'}$.

Let us show that this evaluation can be taken uniform in $L' \leqslant \ell$. The problem comes from the evaluation of the number of points at distance at most $\ell - L' + 4\delta$ from y by $(2m)^{g(\ell-L'+4\delta)}$: when $\ell - L' + 4\delta$ is not large enough, this cannot be taken uniform. So take some $\varepsilon > 0$ and first suppose that $L' \leqslant (1-\varepsilon)\ell$, so that $\ell - L' + 4\delta \geqslant \varepsilon'\ell$ for some $\varepsilon' > 0$. The evaluation of the number of points at distance at most $\ell - L' + 4\delta$ from y by $(2m)^{g(\ell-L'+4\delta)}$ can thus be taken uniform in L' in this interval.

Second, let us suppose that $L' \geqslant (1-\varepsilon)\ell$. Apply the trivial estimate that the number of points at distance $\ell - L' + 4\delta \leqslant \varepsilon\ell + 4\delta$ from y is less than $(2m)^{\varepsilon\ell+4\delta}$. The probability that X_ℓ is equal to one of them is roughly less than $(2m)^{\varepsilon\ell-g\ell} \leqslant (2m)^{-(g-\varepsilon)L'}$ uniformly for these values of L'.

So for any ε, we can show that for any $L' \leqslant \ell$, the probability at play is uniformly roughly less than $(2m)^{-(g-\varepsilon)L'}$. Writing out the definition shows that this exacly says that our probability is less than $(2m)^{-gL'}$ uniformly in L'.

Now summing for L' between L and ℓ introduces only a polynomial factor in ℓ. $\square$

Corollary 21 – *Let $x \in G$ and $L \leqslant 2\ell$. Then*

$$\Pr\left(\|xX_\ell\| \leqslant \|x\| + \ell - L\right) \lesssim (2m)^{-gL/2}$$

and

$$\Pr((\|X_\ell x\| \leqslant \|x\| + \ell - L) \lesssim (2m)^{-gL/2}$$

uniformly in L.

Proof – Note that the second case follows from the first one applied to x^{-1} and X_ℓ^{-1}, and symmetry of the law of X_ℓ.

For the first case, apply Proposition 20 to X_ℓ and x^{-1} and write out the definition of the Gromov product. $\square$

Proposition 22 – *For any fixed N, for any $x \in G$, uniformly for any $L \leqslant 2\ell$ we have*

$$\Pr\left(\left\|xX_\ell^N\right\| \leqslant \|x\| + \ell - L\right) \lesssim (2m)^{-gL/2}$$

and

$$\Pr\left(\left\|X_\ell^N x\right\| \leqslant \|x\| + \ell - L\right) \lesssim (2m)^{-gL/2}$$

Proof – Again, note that the second inequality follows from the first one by taking inverses and using symmetry of the law of X_ℓ^N.

Suppose $\left\|xX_\ell^N\right\| \leqslant \|x\|+\ell-L$. Let $x_1, x_2, \ldots, x_N$ be N random uniformly chosen geodesic words of length ℓ/N. Let $L_i \leqslant 2\ell/N$ such that $\|xx_1 \ldots x_i\| = \|xx_1 \ldots x_{i-1}\| + \ell/N - L_i$. By N applications of Corollary 21, the probability of such an event is roughly less than $(2m)^{-g\varepsilon \sum L_i/2}$. But·we have $\sum L_i \geqslant L$. Now the number of choices for the L_i's is at most $(2\ell)^N$, which is polynomial in ℓ, hence the proposition. $\square$

Of course, this is not uniform in N.

We now turn to satisfaction of Axioms 3 and 4' (1 and 2 being trivially satisfied). We work under the model of X_ℓ^N. Let x be a subword of X_ℓ^N. By taking N large enough (depending on $|x|/\ell$), we can suppose that x begins and ends on a multiple of ℓ/N. If not, throw away an initial and final subword of x of length at most ℓ/N. In the estimates, this will change $\|x\|$ in $\|x\| - 2\ell/N$ and, if the estimate is of the form $(2m)^{-\beta\|x\|}$, for each $\varepsilon > 0$ we can find an N such that we can prove the estimate $(2m)^{-\beta(1-\varepsilon)\|x\|}$.

Now if something is roughly less than $(2m)^{-\beta(1-\varepsilon)\|x\|}$ for every $\varepsilon > 0$, it is by definition roughly less than $(2m)^{-\beta\|x\|}$.

Note that taking N depending on the relative length of the subword is correct since we did not ask the estimates to be uniform in this length.

The main advantage of this model is that now, the law of a subword is independent of the law of the rest of the word, so we do not have to care about the conditional probabilities in the axioms.

Proposition 23 – *Axiom 3 is satisfied for random geodesic words, with exponent $g/2$.*

Proof – Let x and y be subwords. The word x is a product of $N\,|x|\,/\ell$ geodesic words of length ℓ/N, and the same holds for y. Now take two fixed words u, v, and let us evaluate the probability that $xuyv = e$.

Fix some $L \leqslant \ell$, and suppose $\|x\| = L$. By Proposition 22 starting at e, this occurs with probability $(2m)^{-g(|x|-L)/2}$. Now we have $\|xu\| \geqslant L - \|u\|$, but $\|xuy\| = \|v^{-1}\|$. By Proposition 22 starting at xu this occurs with probability $(2m)^{-g(L-\|u\|+|y|-\|v\|)/2}$.

So the total probability is at most the number of choices for u times the number of choices for L times $(2m)^{-g(|x|-L)/2}$ times $(2m)^{-g(L-\|u\|+|y|-\|v\|)/2}$. Hence the proposition. □

Proposition 24 – *Axiom 4' is satisfied for random geodesic words, with exponent $g/2$.*

Proof – Taking notations as in the definitions, let x be a subword of X_ℓ^N of length $b\ell$ with $b \leqslant 1$. The law of x is $X_{b\ell}^{bN}$.

Note that applying Proposition 22 starting with the neutral element e shows that $\Pr(\|x\| \leqslant L) \lesssim (2m)^{-g(|x|-L)/2}$.

Fix a u of length at most n and consider a cyclic conjugate y of xu.

First, suppose that the cutting made in xu to get the cyclic conjugate y was made in u, so that $y = u''xu'$ with $u = u'u''$. In this case, we have $\|y\| \geqslant \|x\| - \|u''\| - \|u\| \geqslant \|x\| - |u|$, and so we have $\Pr(\|y\| \leqslant C\log\ell) \leqslant \Pr(\|x\| \leqslant C\log\ell + \|u\|) \lesssim (2m)^{-g(|x|-C\log\ell-|u|)/2} \approx (2m)^{g|u|/2-g|x|/2}$.

Second, suppose that the cutting was made in x, so that $y = x''ux'$ with $x = x'x''$.

Up to small words of length at most ℓ/N at the beginning and end of x, the words x' and x'' are products of randomly chosen geodesic words of length ℓ/N.

Apply Proposition 22 starting with the element u, multiplying on the right by x', then on the left by x''. This shows that $\Pr(\|y\| \leqslant \|u\|+|x'|+|x''|-L) \lesssim (2m)^{-gL/2}$, hence the evaluation, taking $L = |x'| + |x''| + \|u\| - C\log\ell$.

To conclude, observe that there are at most $(2m)^{|u|}$ choices for u and at most $|x| + |u|$ choices for the cyclic conjugate, hence an exponential factor in $|u|$. □

5.1.3 The case of random reduced words

Recall η is the cogrowth of the group G, i.e. the number of reduced words of length ℓ which are equal to e is roughly $(2m-1)^{\eta\ell}$.

Proposition 25 – *Axioms 1, 2, 3, 4' are satisfied by random uniformly chosen reduced words, or random uniformly chosen cyclically reduced words, with exponent $1-\eta$ (in base $2m-1$).*

The proof follows essentially the same lines as that for plain random words. We do not include it explicitly here.

Nevertheless, there are two changes encountered.

The first problem is that we do not have as much independence for reduced words as for plain words. Namely, the occurrence of a generator at position i prevents the occurrence of its inverse at position $i+1$.

We solve this problem by noting that, though the $(i+1)$-th letter depends on what happened before, the $(i+2)$-th letter does not depend too much.

Indeed, say the i-th letter is x_j. Now it is immediate to check that the $(i+2)$-th letter is x_j with probability $1/(2m-1)$, and is each other letter with probability $(2m-2)/(2m-1)^2$. This is close to a uniform distribution up to a factor of $(2m-2)/(2m-1)$.

This means that, conditioned by the word up to the i-th letter, the law of the word read after the $(i+2)$-th letter is, up to a constant factor, an independently chosen random reduced word.

This is enough to allow to prove satisfaction of the axioms for random reduced words by following the same lines as for plain random words.

The second point to note is that a reduced word is not necessarily cyclically reduced. The end of a reduced word may collapse with the beginning. Collapsing along L letters has probability precisely $(2m-1)^{-L}$, and the induced length loss is $2L$. So this introduces an exponent $1/2$, but the cogrowth η is bigger than $1/2$ anyway.

Note that everything works equally fine with reduced and cyclically reduced words (the difference being non-local), with the same critical density $1-\eta$.

5.2 Triviality of the quotient in large density

Recall G is a hyperbolic group generated by $S = a_1^{\pm 1}, \ldots, a_m^{\pm 1}$. Let R be a set of $(2m)^{d\ell}$ randomly chosen words of length ℓ. We will study $G/\langle R\rangle$.

As was said before, because triviality of the quotient involves small-scale phenomena, we have to work separately on plain random words, reduced random words or random geodesic words.

Generally speaking, the triviality of the quotient reduces essentially to the following fact, which is analogue to the fact that two (say generic projective complex algebraic) submanifolds whose sum of dimensions is bigger than the ambient dimension do intersect (cf. our discussion of the model of random groups in the introduction).

Basic intersection theory for random sets – *Let S be a set of N elements. Let α, β be two numbers in $[0;1]$ such that $\alpha+\beta>1$. Let A be a given part of S of cardinal N^α. Let B be a set of N^β randomly uniformly chosen elements of S. Then $A \cap B \neq \varnothing$ with probability tending to 1 as $N \to \infty$ (and the intersection is arbitrarily large with growing N).*

This is of course a variation on the probabilistic pigeon-hole principle where $A = B$.

Remark – Nothing in what follows is specific to quotients of hyperbolic groups: for the triviality of a random quotient by too many relators, any group would do.

5.2.1 The case of plain random words

We suppose that $d > 1 - \theta$.

Recall that θ is the gross cogrowth of the group, i.e. that

$$\theta = \lim_{\ell\to\infty, \ell \text{ even}} \frac{1}{\ell} \log_{2m} \#\{w \in B^\ell, w = e \text{ in } G\}$$

We want to show that the random quotient $G/\langle R\rangle$ is either $\{1\}$ or $\mathbb{Z}/2\mathbb{Z}$. Of course the case $\mathbb{Z}/2\mathbb{Z}$ occurs when ℓ is even and when the presentation of G does not contain any odd-length relation.

To use gross cogrowth, we have to distinguish according to parity of ℓ. We will treat only the least simple case when ℓ is even. The other case is even simpler.

Rely on the intersection theory for random sets stated above. Take for A the set of all words of length $\ell - 2$ which are equal to e in G. There are roughly $(2m)^{\theta(\ell-2)} \approx (2m)^{\theta\ell}$ of them. Take for B the set made of the random words of R with the last two letters removed, and recall that R consists of $(2m)^{d\ell}$ randomly chosen words with $d > 1 - \theta$.

Apply the intersection principle: very probably, these sets will intersect. This means that in R, there will probably be a word of the form wab such that w is trivial in G and a, b are letters in S or S^{-1}.

This means that in the quotient $G/\langle R\rangle$, we have $ab = e$.

Now as $d+\theta > 1$ this situation occurs arbitrarily many times as $\ell \to \infty$. Due to our uniform choice of random words, the a and b above will exhaust all pairs of generators of S and S^{-1}.

Thus, in the quotient, the product of any two generators $a, b \in S \cup S^{-1}$ is equal to e. Hence the quotient is either trivial or $\mathbb{Z}/2\mathbb{Z}$ (and is it trivial as soon as ℓ is odd or the presentation of G contains odd-length relators).

This proves the second part of Theorem 4.

5.2.2 The case of random geodesic words

When taking a random quotient by geodesic words of the same length, some local phenomena may occur. For example, the quotient of $\mathbb{Z}$ by any number of randomly chosen elements of norm ℓ will be $\mathbb{Z}/\ell\mathbb{Z}$. Think of the occurrence of either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$ in a quotient by randomly chosen non-geodesic words.

In order to avoid this phenomenon, we will consider a random quotient by randomly chosen elements of norm comprised between $\ell - L$ and $\ell + L$ for some fixed small L. Actually we will take $L = 1$.

Recall g is the growth of the group, that is, the number of elements of norm ℓ is roughly $(2m)^{g\ell}$.

We now prove that a random quotient of any group G by $(2m)^{d\ell}$ randomly chosen elements of norm $\ell - 1$, ℓ and $\ell + 1$, with $d > g/2$, is trivial with probability tending to 1 as $\ell \to \infty$.

(By taking $(2m)^{d\ell}$ elements of norm ℓ, $\ell + 1$ or $\ell - 1$ we mean either taking $(2m)^{d\ell}$ elements of each of these norms, or taking 1/3 at each length, or deciding for each element with a given positive probability what its norm will be, or any other roughly equivalent scheme.)

Let a be any of the generators of the group. Let x be any element of norm ℓ. The product xa is either of norm ℓ, $\ell + 1$ or $\ell - 1$.

Let S be the sphere of radius ℓ, we have $|S| \approx (2m)^{g\ell}$.

Let R be the set of random words taken. Taking $d > g/2$ precisely amounts to taking more than $|S|^{1/2}$ elements of S.

Let R' be the image of R by $x \mapsto xa$. By an easy variation on the probabilistic pigeon-hole principle applied to R, there will very probably be one element of R lying in R'. This means that R will contain elements x and y such that $xa = y$. Hence, $a = e$ in the quotient by R.

As this will occur for any generator, the quotient is trivial. This proves the second part of Theorem 3.

5.2.3 The case of random reduced words

For a quotient by random reduced words in density $d > 1 - \eta$ (where η is the cogrowth of the group), the proof of triviality is nearly identical to the case of a quotient by plain random words, except that inorder to have the number of words taken go to infinity, we have to suppose that G is non-elementary.

5.3 Elimination of the virtual centre

Theorem 9 only applies to random quotients of hyperbolic groups with trivial virtual centre. In order not to get this restriction in our applications, we have to show that this assumption is harmless. The way to do this is simply to quotient by the virtual centre; but, for example, geodesic words in the quotient are not geodesic words in the original group, and moreover, the growth, cogrowth and gross cogrowth may be different. Thus something should be said.

Recall the virtual centre of a hyperbolic group is the set of elements whose action on the boundary at infinity is trivial. It is a normal subgroup (as it is defined as the kernel of some action). It is finite, as any element of the virtual centre has force 1 at each point of the boundary, and in a (non-elementary) hyperbolic group, the number of elements having force less than a given constant at some point is finite (cf. [GH], p. 155). See [Ch3] for an exposition of basic properties and to get an idea of the kind of problems arising because of the virtual centre.

Let H be the virtual centre of G and set $G' = G/H$. The quotient G' has no virtual centre.

5.3.1 The case of plain or reduced random words

Note that the set R is the same, since the notion of plain random word or random reduced word is defined independently of G or G'.

As $(G/H)/\langle R\rangle = (G/\langle R\rangle)/H$, and as a quotient by a finite normal subgroup is a quasi-isometry, $G/\langle R\rangle$ will be infinite hyperbolic if and only if $G'/\langle R\rangle$ is.

So in order to prove that we can assume a trivial virtual centre, it is enough to check that G and G/H have the same cogrowth and gross cogrowth, so that the notion of a random quotient is really the same.

We prove it for plain random words, as the case of reduced words is identical with θ replaced with η and $2m$ replaced with $2m-1$.

Proposition 26 – *Let H be a subset of G, and n an integer. Then*

$$\Pr(\exists u \in G, |u| = n, B_\ell u \in H) \leqslant (2m)^n \Pr(B_{\ell+n} \in H)$$

Proof – Let H_n be the n-neighborhood of H in G. We have that $\Pr(B_{\ell+n} \in H) \geqslant (2m)^{-n} \Pr(B_\ell \in H_n)$. □

Corollary 27 – *A quotient of a group by a finite normal subgroup has the same gross cogrowth.*

Proof – Let H be a finite subgroup of G and let $n = \max\{\|h\|, h \in H\}$ so that H is included in the n-neighborhood of e. Then $\Pr(B_\ell =_{G/H} e) = \Pr(B_\ell \in H) \leqslant \sum_{k\leqslant n} (2m)^k \Pr(B_{\ell+k} = e) \lesssim (2m)^{-(1-\theta)\ell}$. □

Remark – Gross cogrowth is the same only if defined with respect to the same set of generators. For example, $F_2 \times \mathbb{Z}/2\mathbb{Z}$ presented by a, b, c with $ac = ca$, $bc = cb$ and $c^2 = e$ has the same gross cogrowth as F_2 presented by a, b, c with $c = e$.

So in this case, we can safely assume that the virtual centre of G is trivial.

5.3.2 The case of random geodesic words

A quotient by a finite normal subgroup preserves growth, so G and G' have the same growth.

But now a problem arises, as the notion of a random element of norm ℓ differs in G and G'. So our random set R is not defined the same way for G and G'.

Let us study the image of the uniform measure on the ℓ-sphere of G into G'. Let L be the maximal norm of an element in H. The image of this sphere is contained in the spheres of radius between $\ell - L$ and $\ell + L$.

The map $G \to G'$ is of index $|H|$. This proves that the image of the uniform probability measure μ_ℓ on the sphere of radius ℓ in G is, as a measure, at most $|H|$ times the sum of the uniform probability measures on the spheres of G' of radius between $\ell - L$ and $\ell + L$. In other words, it is roughly less than the uniform probability measure ν_ℓ on these spheres.

The uniform measure ν_ℓ on the spheres of radius between $\ell - L$ and $\ell + L$ (for a fixed L) satisfies our axioms, so does the image in G' of the uniform measure on the sphere of radius ℓ in G.

So we can apply Theorem 9 to the quotient of G' by a set R' of random words chosen using measure ν_ℓ. This random quotient will be non-elementary hyperbolic for $d < g/2$.

By Remark 10, for a random set R picked from measure μ_ℓ (the one we are interested in), the quotient $G'/\langle R\rangle$ will be non-elementary hyperbolic as well.

But $G'/\langle R\rangle = G/H/\langle R\rangle = G/\langle R\rangle/H$, and quotienting $G/\langle R\rangle$ by the finite normal subgroup H is a quasi-isometry, so $G/\langle R\rangle$ is non-elementary hyperbolic if and only if $G'/\langle R\rangle$ is.

6 Proof of the main theorem

We now proceed to the proof of Theorem 9.

G is a hyperbolic group without virtual centre generated by $S = a_1^{\pm 1}, \ldots, a_m^{\pm 1}$. Say that G has presentation $\langle S; Q\rangle$. Let R be a set of random words of density at most d picked under the measure μ_ℓ. We will study $G/\langle R\rangle$.

Let $\beta = \min(\beta_2, \beta_3, \beta_4)$ where $\beta_2, \beta_3, \beta_4$ are given by the axioms. We assume that $d < \beta$.

We will study Dehn diagrams in the group $G/\langle R\rangle$. If G is presented by $\langle S; Q\rangle$, call *old relator* an element of Q and *new relator* an element of R.

We want to show that Dehn diagrams of $G/\langle R\rangle$ satisfy a linear isoperimetric inequality. Let D be such a diagram. D is made of old and new relators. Denote by D' the subdiagram of D made of old relators and by D'' the subdiagram of D made of new relators.

If $\beta = 0$ there is nothing to prove. Hence we suppose that $\beta > 0$. In the examples we consider, this is equivalent to G being non-elementary (that is, neither finite nor quasi-isometric to $\mathbb{Z}$).

6.1 On the lengths of the relators

In order not to make the already complex notations even heavier, *we will suppose that all the words taken from μ_ℓ are of length ℓ.* So R is made of $(2m)^{d\ell}$ words of length ℓ. This is the case in all the applications given in this text.

Indeed, there are only three steps of the proof in which the length of the elements matters:

1. As we are to apply asymptotic estimates, the length of the elements must tend to infinity.

2. The hyperbolic local-to-global theorem of Appendix A crucially needs that the ratio of the lengths of relators be bounded independently of ℓ.

3. In our probability estimates, the number of possible lengths for a given relator must be subexponential in ℓ.

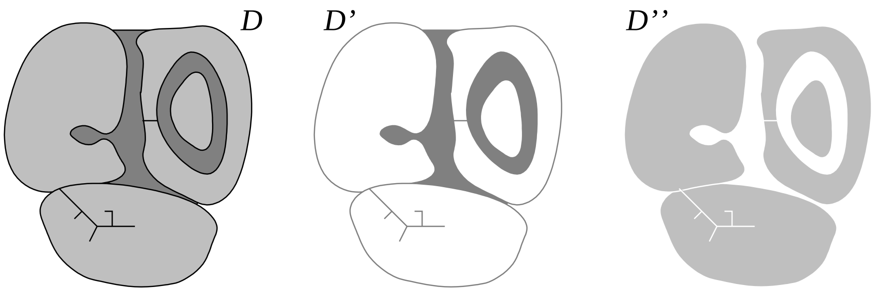
All these properties are guaranteed by Axiom 1.

6.2 Combinatorics of Dehn diagrams of the quotient

We now proceed to the application of the program outlined in section 3. We suggest that the reader re-read this section now.

We consider a Dehn diagram D of $G/\langle R\rangle$. Let D' be the part of D made of old relators of the presentation of G, and D'' the part made of new relators in R.

Redefine D' by adding to it all edges of D'': this amounts to adding some filaments to D'. This way, we ensure that faces of D'' are isolated and that D' is connected; and that if a face of D'' lies on the boundary of D, we have a filament in D', such that D'' does not intersect the boundary of D; and last, that if the diagram D'' is not regular (see section 1 for definition), we have a corresponding filament in D'.



This manipulation ensures that each face of D'' is in contact only with D', and that we only have to deal with relations between old and new relators.

We want to show that if D is minimal, then it satisfies some isoperimetric inequality. In fact, as in the case of random quotients of a free group, we do not really need that D is minimal. We need that D is reduced in a slighlty stronger meaning than previously, which we define now.

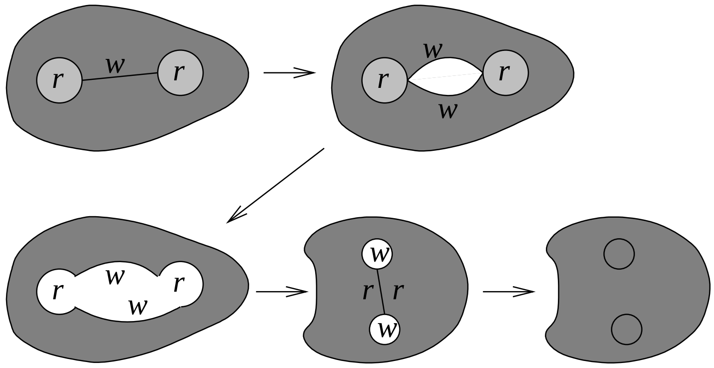
Definition 28 – *A Dehn diagram $D = D' \cup D''$ on $G/\langle R \rangle$ (with D' and D'' as above) is said to be* strongly reduced *with respect to G if there is no pair of faces of D'' bearing the same relator with opposite orientations, such that their marked starting points are joined in D' by a path representing the trivial element in G.*

In particular, a strongly reduced diagram is reduced.

Proposition 29 – *Every Dehn diagram has a strong reduction, that is, there exists a strongly reduced diagram with the same boundary.*

In particular, to ensure hyperbolicity of a group it is enough to prove the isoperimetric inequality for all strongly reduced diagrams.

Proof – Suppose that some new relator r of D'' is joined to some r^{-1} by a path w in D' representing the trivial element in G. Then incise the diagram along w and apply surgery to cancel r with r^{-1}. This leaves a new diagram with two holes w, w^{-1}. Simply fill up these two holes with diagrams in G bordered by w (this is possible precisely since w is the trivial element of G).



Note that this way we introduce only old relators and no new ones in the diagram. Iterate the process to get rid of all annoying pairs of new relators. □

We will show that any strongly reduced Dehn diagram D such that D' is minimal very probably satisfies some linear isoperimetric inequality. By the local-to-global principle for hyperbolic spaces (Cartan-Hadamard-Gromov-Papasoglu theorem, cf. Appendix A), it is enough to show it for diagrams having less than some fixed number of faces. More precisely, we will show the following.

PROPOSITION 30 – *There exist constants $\alpha, \alpha' > 0$ (depending on G and d but not on ℓ) such that for any integer K, for any Dehn diagram $D = D' \cup D''$ satisfying the following conditions:*

- *D' is minimal among Dehn diagrams in G with the same boundary;*
- *The number of faces of D'' is at most K;*
- *D is strongly reduced with respect to G;*

then D satisfies the isoperimetric inequality

$$|\partial D| \geqslant \alpha\ell\,|D''| + \alpha'\,|D'|$$

with probability exponentially close to 1 as $\ell \to \infty$.

(Of course, the constant implied in "exponentially close" depends on K.)

Before proceeding to the proof of this proposition, let us see how it implies hyperbolicity of the group $G/\langle R\rangle$, as well as that of all intermediate quotients.

PROPOSITION 31 – *A group satisfying the conditions of Proposition 30 is hyperbolic. Better, there exist constants $\alpha_1, \alpha_2 > 0$ such that for any strongly reduced diagram D such that D' is minimal, we have*

$$|\partial D| \geqslant \alpha_1\ell\,|D''| + \alpha_2\,|D'|$$

REMARK 32 – Proposition 31 implies that a quotient of G by a smaller set $R' \subset R$ is hyperbolic as well. Indeed, any strongly reduced diagram on R' is, in particular, a strongly reduced diagram on R.

PROOF – By our strongly reduction process, for any Dehn diagram there exists another Dehn diagram D with the same boundary, such that D' is minimal (otherwise replace it by a minimal diagram with the same boundary) and D is strongly reduced. Thus, it is enough to show the isoperimetric inequality for strongly reduced diagrams to ensure hyperbolicity.

We want to apply Proposition 41. Take for property P in this proposition "to be strongly reduced". Recall the notations of the appendix: $L_c(D) = |\partial D|$ is the boundary length of D, and $A_c(D)$ is the area of D in the sense that a relator of length L has area L^2. Note that $\ell\,|D''| + |D'| \geqslant A_c(D)/\ell$.

Take a Dehn diagram D such that $k^2/4 \leqslant A_d(D) \leqslant 480k^2$ for some $k^2 = K\ell^2$ where K is some constant independant of ℓ to be chosen later. As $A_d(D) \leqslant K\ell^2$, we have $|D''| \leqslant K$. Proposition 30 for this K tells us that $L_c(D) = |\partial D| \geqslant \alpha\ell\,|D''| + \alpha'\,|D'| \geqslant \min(\alpha, \alpha')A_c(D)/\ell$. Thus

$$L_c(D)^2 \geqslant \min(\alpha, \alpha')^2 A_c(D)^2/\ell^2 \geqslant \min(\alpha, \alpha')^2 A_c(D)K/4$$

as $A_c(D) \geqslant k^2/4$, so taking $K = 10^{15}/\min(\alpha,\alpha')^2$ is enough to ensure that the conditions of Proposition 41 are fulfilled by $K\ell^2$. (The important point is that this K is independent of ℓ.)

The conclusion is that any strongly reduced Dehn diagram D satisfies the linear isoperimetric inequality

$$L_c(D) \geqslant A_c(D)\min(\alpha,\alpha')/10^{12}\ell$$

and, fiddling with the constants and using the isoperimetry from D, we can even put it in the form

$$|\partial D| \geqslant \alpha_1 \ell\, |D''| + \alpha_2\, |D'|$$

if it pleases, where $\alpha_{1,2}$ depend on G and d but not on ℓ.

So the proposition above, combined with the local-to-global hyperbolicity principle of Appendix A, is sufficient to ensure hyperbolicity. □

A glance through the proof can even show that if ℓ is taken large enough, the constant α_2 in the inequality

$$|\partial D| \geqslant \alpha_1 \ell\, |D''| + \alpha_2\, |D'|$$

is arbitrarily close to the original isoperimetry constant in G.

This suggests, in the spirit of [Gro4], to iterate the operation of taking a random quotient, at different lengths ℓ_1, then ℓ_2, etc., with fast growing ℓ_i. The limit group will not be hyperbolic (it will be infinitely presented), but it will satisfy an isoperimetric inequality like

$$|\partial D| \geqslant \alpha \sum_{f \text{ face of } D} \ell(f)$$

where $\ell(f)$ denotes the length of a face. This property could be taken as a definition of a kind of loose hyperbolicity, which should be related in some way to the notion of "fractal hyperbolicity" proposed in [Gro4].

Now for the proof of Proposition 30.

We have to assume that D' is minimal, otherwise we know nothing about its isoperimetry in G. But as in the case of a random quotient of F_m (section 2), the isoperimetric inequality will not only be valid for minimal diagrams but for all (strongly reduced) configurations of the random relators.

If $D'' = \varnothing$ then $D = D'$ is a Dehn diagram of G and as D' is minimal, it satifies the inequality $|\partial D| \geqslant C\,|D|$ as this is the isoperimetric inequality in G. So we can take $\alpha' = C$ and any α in this case.

Suppose that the old relators are much more numerous than the new ones, more precisely that $|D'| \geqslant 4\,|D''|\,\ell/C$. In this case as well, isoperimetry in G is enough to ensure isoperimetry of D. Note that D' is a diagram with at most $|D''|$ holes. We have of course that $|\partial D| \geqslant |\partial D'| - |\partial D''| \geqslant |\partial D'| - |D''|\,\ell$.

By Proposition 52 for diagrams with holes in G, we have that $|\partial D'| \geqslant C|D'| - |D''|\lambda(2+4\alpha\log|D'|)$. So, for ℓ big enough,

$$\begin{aligned}
|\partial D| &\geqslant |\partial D'| - |D''|\,\ell \\
&\geqslant C|D'| - |D''|\,\ell - |D''|\,\lambda(2+4\alpha\log|D'|) \\
&\geqslant C|D'|/3 + (C|D'|/3 - |D''|\,\ell) \\
&\quad + (C|D'|/3 - |D''|\,\lambda(2+4\alpha\log|D'|)) \\
&\geqslant C|D'|/3 + (4|D''|\,\ell/3 - |D''|\,\ell) \\
&\quad + (4|D''|\,\ell/3 - |D''|\,\lambda(2+4\alpha\log 4|D''|\,\ell/C)) \\
&\geqslant C|D'|/3 + \ell|D''|/3
\end{aligned}$$

as for ℓ big enough, the third term is positive. So in this case we can take $\alpha = 1/3$ and $\alpha' = C/3$.

So *we now suppose that* $1 \leqslant |D''| \leqslant K$ *and that* $|D'| \leqslant 4|D''|\,\ell/C$.

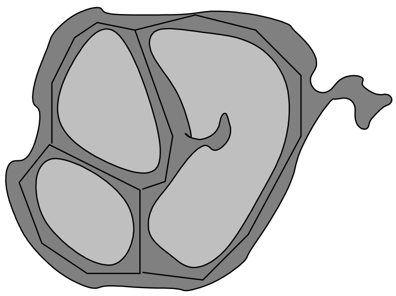
6.3 Coarsening of a Dehn diagram

We now define the coarsening of a Dehn diagram: this will be the Dehn diagram "seen at the scale of the new relators of R". We use the fact that D' is very narrow (at the scale of ℓ), so that at this scale D looks like a Dehn diagram with respect to the new relators, with some narrow "glue" (that is, old relators) between faces.

The diagram D' has at most K holes. After Corollary 53, it is $\lceil\alpha\log|D'|\rceil + K(4\lceil\alpha\log|D'|\rceil + 2)$-narrow. As $|D'| \leqslant 4K\ell/C$, this is less than $E\log\ell$ for some constant E depending on G and K but not on ℓ.

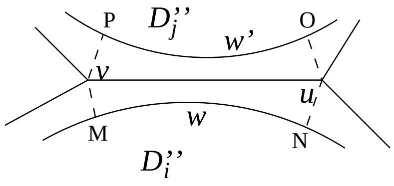
So D' is $E\log\ell$-narrow. This means that a point of D' is either $E\log\ell$-close to some point of D'' or to some point of the boundary of D.

It is therefore possible to partition D into (at most) $K+1$ subcomplexes $D_1, \dots, D_{K+1}$ such that D_i ($i \leqslant K$) is included in the $E\log\ell$-neighborhood of the i-th face of D'', and D_{K+1} is included in the $E\log\ell$-neighborhood of the boundary. The partition can be taken to be made of topological disks (except for D_{K+1} which is an annulus; say we simply cut it into two pieces).

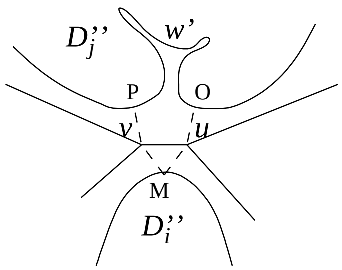


The D_i's for $1 \leqslant i \leqslant K$ form a planar graph X, which is a kind of Dehn diagram at the scale of the new relators. Denote by D''_i the i-th face of D'', so that $D''_i \subset D_i$.

Each internal edge of X defines a word in the following way. Say that the internal edge f in X lies between faces D_i and D_j. Consider the two endpoints x, y of f. By construction, these endpoints are $E \log \ell$-close to D''_i and D''_j. Let M be a point of the boundary of D''_i which is $E \log \ell$-close to x, and define similarly N on D''_i close to y, O on D''_j close to y and P'' on D_j close to x. Now the quadrilateral $MNOP$ is bordered by a word $wuw'v$ such that w lies on the border of D''_i, w' lies on the border of D''_j, and u and v are words of length at most $2E \log \ell$.



As there can be invaginations of D' into D'', the lengths of w and w' may not be equal at all. It may even be the case that one of these two words is of length 0, as in the following picture. This is not overmuch disturbing but should be kept in mind.



Similarly, every external edge of X defines a word $bub'v$ with b lying on the boundary of some D''_i, with b' lying on the boundary of the whole diagram D, and u, v of length at most $2E \log \ell$.

Now we begin to define the *coarsening* $\overline{X}$ of D (there will still be some more decoration added to it below). This is basically the graph X with some decoration on it. Namely, take the graph X. Each face of it is a face of D'', that is, a relator in R with an orientation and a starting point. Put on each face of X a number between 1 and K so that two faces corresponding to the same relator of X get the same number. Also mark the orientation and starting point. Also mark on each internal edge of X, the lengths of the two words w, w' defined above (each associated to one of the two faces bordered by the edge). Also mark on each external edge, the length of the word b defined above (which is a word lying on the boundary of the corresponding face of D'').

So the coarsening $\overline{X}$ closely resembles a daDd, except that each edge bears two lengths instead of one. *From now on, we redefine a daDd to be such a decorated graph.*

A daDd is said to be *fulfillable* if it is the coarsening of some *strongly reduced* Dehn diagram D of $G/\langle R\rangle$. We have to show that any fulfillable daDd satisfies some linear isoperimetric inequality with high probability.

Note that as $\overline{X}$ is a planar graph with at most K faces and each vertex of which has multiplicity at least 3 (by construction), by the Euler formula the number of edges of X is at most $3K$.

6.4 Graph associated to a decorated abstract Dehn diagram

As in the case of random quotients of the free group, we will construct an auxiliary graph Γ summarizing all conditions imposed by a daDd on the random relators of R. But instead of imposing equality between letters of these relators, the conditions will rather be interpreted as equality modulo G.

Let now D be a daDd. We will evaluate the probability that it is fulfillable by the relators of R.

Each face of D bears a number between 1 and $|D|$. Let n be the number of such distinct numbers, we have $n \leqslant |D|$. Suppose for the sake of simplicity that these n distinct numbers are $1, 2, \ldots, n$.

To fulfill the diagram is to give n relators $r_1, \ldots, r_n$ satisfying the conditions that if we put these relators in the corresponding faces, and if we "thicken" the edges of D by words representing the identity in G, then we get a (strongly reduced) Dehn diagram of $G/\langle R\rangle$.

We now construct the auxiliary graph Γ.

Take $n\ell$ points as vertices of Γ, arranged in n parts of ℓ vertices called the *parts* of Γ. Interpret the k-th vertex of the i-th part as the k-th letter of relator r_i in R.

We now explain what to take as edges of Γ.

Let f be an edge of D. Say f is an edge between faces bearing numbers i and i'. The edge f bears two lengths L, L' corresponding to a set of L successive vertices in the i-th part of Γ and to L' successive vertices in the i'-th group of Γ.

Add to Γ a special vertex w called an *internal translator*. Add edges between w and each of the L vertices of the i-th part of Γ represented by edge f; symmetrically, add edges between w and each of the L' vertices of the i'-th part of Γ.

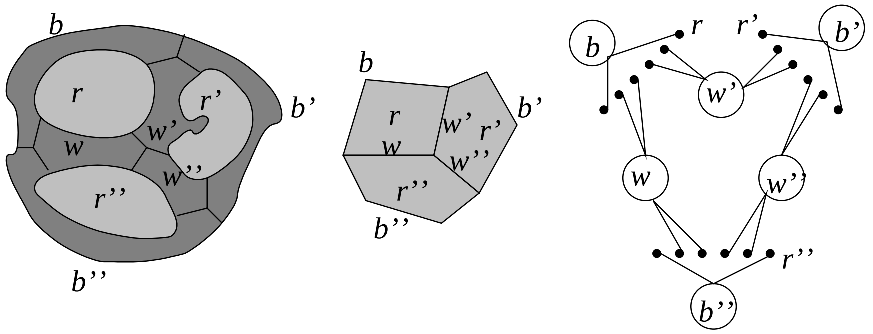
(This may result in double edges if $i = i'$; we will deal with this problem later.)

Follow this process for all internal edges of D. After this construction, there are as many translators as internal edges of D. Each translator is connected with two (or maybe one if $i = i'$) parts of Γ. The number of edges of Γ is the sum of all the lengths bore by internal edges of D.

As two faces of D can bear the same number (the same relator of R), a vertex of Γ is not necessarily of multiplicity one. The multiplicity of a vertex of the i-th part is at most the number of times relator i appears on a 2-face of D.

For each external edge of D (say adjacent to face i, bearing length L), add a special vertex b to Γ, called a *boundary translator*. Add L edges between b and the L vertices of the i-th part of Γ corresponding to the external edge of D at play.

Here is an example of a simple Dehn diagram on $G/\langle R\rangle$, its coarsening $\overline{X}$, and the associated graph Γ.



As the number of edges of $\overline{X}$ is at most $3K$, the number of internal and boundary translators in Γ is at most $3K$.

Note that each translator corresponds to a word in the Dehn diagram which is equal to e in G.

Indeed, fulfillability of the daDd implies that for each translator in Γ, we can find a word w which is equal to e in G, and such that $w = w_1 u w_2 v$ where u and v are short (of length at most $E \log \ell$) and that w_1 and w_2 are the subwords of the relators of R to which the translator is joined. In the case of random quotients of F_m, we had the relators of R directly connected to each other, imposing equality of the corresponding subwords; here this equality happens modulo translators that are equal to e in G.

6.5 Elimination of doublets

A *doublet* is a vertex of Γ that is joined to some translator by a double edge. This can occur only if in the coarsening of the Dehn diagram, two adjacent faces bear the same relator.

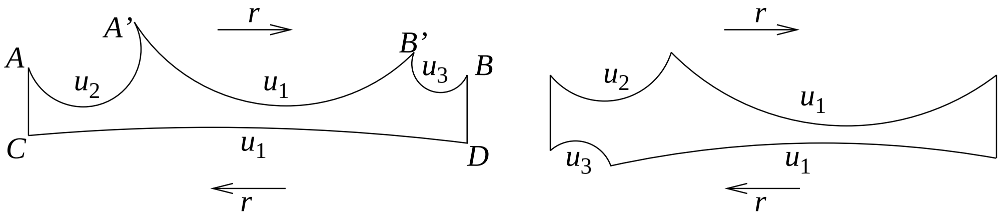
Doublets are annoying since the two sides of the translator are not chosen independently, whereas our argument requires some degree of independence. We will split the corresponding translators to control the occurrences of such a situation.

This section is only technical.

Consider a translator in the Dehn diagram bordered by two faces bearing the same relator r. As a first case, suppose that these two relators are given the same orientation.

Let w be the translator, w writes $w = u\delta_1 u'\delta_2$ where u and u' are subwords of r, and $\delta_{1,2}$ are words of length at most $2E\log\ell$. The action takes place in G. As u and u' need not be geodesic, they do not necessarily have the same length. Let u_1 be the maximum common subword of u and u' (i.e. their intersection as subwords of r). If u_1 is empty there is no doublet.

There are two cases (up to exchanging u and u'): either $u = u_2u_1u_3$ and $u' = u_1$, or $u = u_2u_1$ and $u' = u_1u_3$.



We will only treat the first case, as the second one is similar.

Redefine u_1, u_2 and u_3 to be geodesic words equal to u_1, u_2 and u_3 respectively. In any hyperbolic space, any point on a geodesic joining the two ends of a curve of length L is $(1 + \delta\log L)$-close to that curve (cf. [BH], p. 400). So the new geodesic words are $(1 + \delta\log\ell)$-close to the previous words u_1, u_2, u_3. Hence, up to increasing E a little bit, we can still suppose that D is fulfillable such that D' is $E\log\ell$-narrow, and that u_1, u_2, u_3 are geodesic.

Define points A, A', B, B', C, D as in the figure. The word read while going from A' to B' is the same as that from D to C.

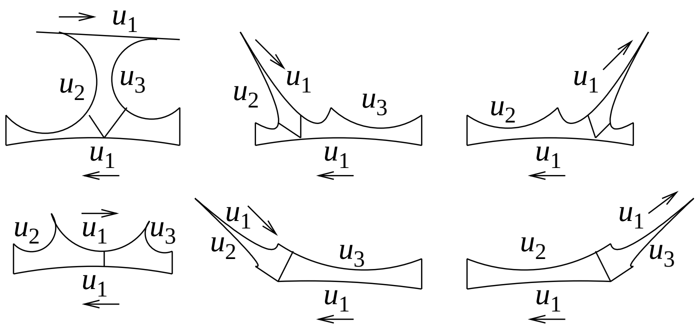
By elementary hyperbolic geometry, and given that the two lateral sides are of length at most $2E\log\ell$, any point on CD is $(2\delta + 2E\log\ell)$-close to some point on AA' or $B'B$, or 2δ-close to some point on $A'B'$.

The idea is to run from D to C, and simultaneously from A' to B' at the same speed. When the two trajectories get $E\log\ell$-close to each other, we cut the translator at this position, and by construction the resulting two parts do not contain any doublets.

Let $L = |u_1|$ and for $0 \leqslant i \leqslant L$, let C_i be the point of DC at distance i from D. Now assign to i a number $\varphi(i)$ between 0 and L as follows: C_i is close to some point C'_i of AB, set $\varphi(C_i) = 0$ if $C'_i \in AA'$, $\varphi(C_i) = L$ if $C'_i \in B'B$, and $\varphi(C_i) = \mathrm{dist}(C'_i, A')$ if $C'_i \in A'B'$.

By elementary hyperbolic geometry (approximation of $A'B'DC$ by a tree), the function $\varphi : [0; L] \to [0; L]$ is decreasing up to 8δ (that is, $i < j$ implies $\varphi(i) > \varphi(j) - 8\delta$). We have $\varphi(0) = L$ and $\varphi(L) = 0$ (up to 8δ). Set i_0 as the smallest i such that $\varphi(i) < i$. This defines a point C_{i_0} on DC and a point C'_{i_0} on $A'B'$ which are 2δ-close.

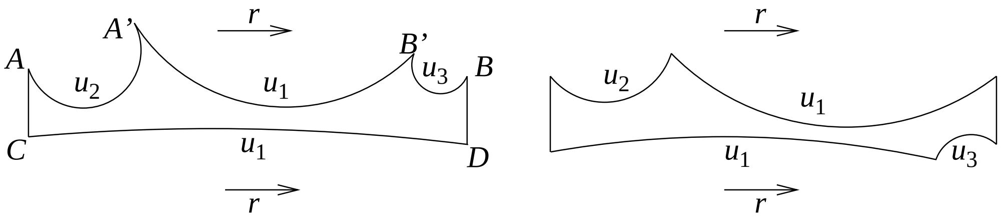
There are six cases depending on whether C'_{i_0} and C'_{i_0-1} belong to AA', $A'B'$ or $B'B$. In each of these cases we can cut the diagram in at most three parts, in such a way that no part contains two copies of some subword of u_1 (except perhaps up to small words of length at most 8δ at the extremities). The cuts to make are from C_{i_0} to C'_{i_0} and/or to C'_{i_0-1}, and are illustrated below in each case.



A translator is a vertex of Γ and by "cutting a translator" we mean that we split this vertex into two, and share the edges according to the figure.

As our second (and more difficult) case, suppose that the translator is bordered by two faces of the diagram bearing the same relator r of R with opposite orientations. This means that the translator w is equal, in G, to $u\delta_1 u'^{-1}\delta_2$ where u and u' are subwords of the relator r, and where $\delta_{1,2}$ are words of length at most $2E\log\ell$.

As above, let u_1 be the maximum common subword of u and u' (i.e. their intersection as subwords of r). There are two cases: $u = u_2u_1u_3$ and $u' = u_1$, or $u = u_2u_1$ and $u' = u_1u_3$.



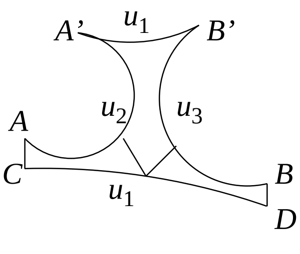
We will only treat the first case, as the second is similar.

As above, redefine u_1, u_2 and u_3 to be geodesic.

Define points A, A', B, B', C, D as in the figure. The word read while going from A' to B' is the same as that from C to D.

By elementary hyperbolic geometry, and given that the two lateral sides are of length at most $2E\log\ell$, any point on CD is $(2\delta + 2E\log\ell)$-close to some point on AA' or $B'B$, or 2δ-close to some point on $A'B'$.

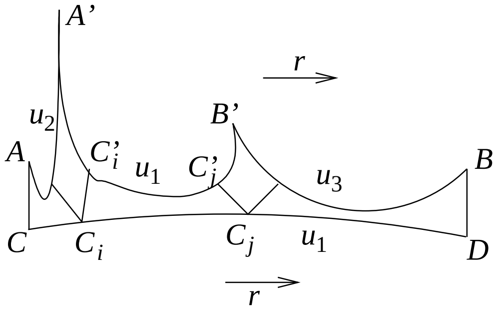
If any point on CD is close to a point on either AA' or BB', we can simply eliminate the doublets by cutting the figure at the last point of CD which is close to AA'. (As above, by cutting the figure we mean that we split the vertex of Γ representing the translator into three new vertices and we share its edges according to the figure.) In this way, we obtain a new graph Γ with the considered doublets removed.



Otherwise, let $L = |u_1|$ and for $0 \leqslant i \leqslant L$, let C_i be the point of CD at distance i from C. Now assign to i a number $\varphi(i)$ between 0 and L as follows: C_i is close to some point C'_i of AB, set $\varphi(C_i) = 0$ if $C'_i \in AA'$, $\varphi(C_i) = L$ if $C'_i \in B'B$, and $\varphi(C_i) = \mathrm{dist}(C'_i, A')$ if $C'_i \in A'B'$.

It follows from elementary hyperbolic geometry (approximation of the quadrilateral $CA'B'D$ by a tree) that $\varphi : [0; L] \to [0; L]$ is an increasing function up to 8δ (that is, $i < j$ implies $\varphi(i) < \varphi(j) + 8\delta$). Moreover, let i be the smallest such that $\varphi(i) > 0$ and j the largest such that $\varphi(j) < L$. Then φ is, up to 8δ, an isometry of $[i; j]$ to $[\varphi(i); \varphi(j)]$ (this is clear on the approximation of $CA'B'D$ by a tree). In other words: the word u_1 is close to a copy of it with some shift $\varphi(i) - i$.

Cut the figure in five: cut between C_i and C'_i, between C_i and a point of AA' close to it, between C_j and C'_j and between C_j and a point of $B'B$ close to it (such points exist by definition of i and j).



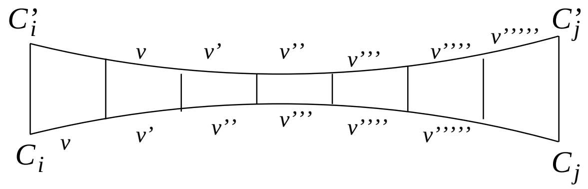
This way, we get a figure in which only the middle part $C_iC_jC'_jC'_i$ of the figure contains two copies of a given piece of u_1. Indeed (from left to right in the figure) the first part contains letters 0 to i of the lower copy of u_1 and no letter of the upper u_1; the second part contains letters 0 to $\varphi(i)$ of the upper u_1 and no letter of the lower u_1; the third part $C_iC_jC'_jC'_i$ contains letters i to j of the lower u_1 and letters $\varphi(i)$ to $\varphi(j)$ of the upper u_1; the fourth and fifth part each contain letters from only one copy of u_1.

First suppose that the intersection of $[i; j]$ and $[\varphi(i); \varphi(j)]$ is empty, or if its size is smaller than $\varepsilon_1 |u_1|$ (for some small ε_1 to be fixed later on, depending on d and G but not on ℓ). Then, in the new graph Γ defined by such cutting of the translator, at most $\varepsilon_1 |u_1|$ of the doublets at play remain. Simply remove these remaining double edges from Γ.

In case the intersection of $[i; j]$ and $[\varphi(i); \varphi(j)]$ is not smaller than $\varepsilon_1 |u_1|$, let us now deal with the middle piece.

Consider the subdiagram $C_iC_jC'_jC'_i$: it is bordered by two subwords u'_1, u''_1 of u_1 of non-empty intersection. The subword u'_1 spans letters i to j of u_1, whereas u''_1 spans letters $\varphi(i)$ to $\varphi(j)$, with $\varphi(j) - \varphi(i) = j - i$ up to 8δ.

First suppose that the shift $\varphi(i) - i$ is bigger than $\varepsilon_2 |u_1|$. Then, chop the figure into sections of size $\varepsilon_2 |u_1|$:



The word read on one side of a section is equal to the word read on the other side of the following section, but there are no more doublets. The original translator has been cut into at most $1/\varepsilon_2$ translators, the length of each of which is at least $\varepsilon_2 |u_1|$.

Second (and last!), suppose that the shift $\varphi(i) - i$ is smaller than $\varepsilon_2 |u_1|$. This means that we have an equality $w_1 v w_2 v^{-1}$ in G, where v is a subword of a random relator r, of length at least $\varepsilon_1 |u_1|$, and with w_1, w_2 words of length at most $\varepsilon_2 |u_1|$.

As the diagram is strongly reduced, w_1 and w_2 are non-trivial in G. As the virtual centre of G has been supposed to be trivial, the probability of this situation is controlled by Axiom 4. Let this translator as is, but mark it (add some decoration to Γ) as being a *commutation translator*. Furthermore, remove from this translator all edges that are not double edges, that is, all edges not corresponding to letters of the v above (there are at most $\varepsilon_1 |u_1|$ of them).

Follow this process for each translator having doublets. After this, some doublets have been removed, and some have been marked as being part of a commutation translator. Note that we suppressed some of the edges of Γ, but the proportion of suppressed edges is less than ε_1 in each translator.

6.6 Pause

Let us sum up the work done so far. Remember the example on page 46.

Proposition 33 – *For each strongly reduced Dehn diagram D of the quotient $G/\langle R\rangle$ such that $|D''| \leqslant K$ and $|D'| \leqslant 4\,|D''|\,\ell/C$, we have constructed a graph Γ enjoying the following properties:*

- Vertices of Γ are of four types: ordinary vertices, internal translators, boundary translators, and commutation translators.
- There are $n\ell$ ordinary vertices of Γ, grouped in n so-called parts, of ℓ vertices each, where n is the number of different relators of R that are present in D. Hence each ordinary vertex of Γ corresponds to some letter of a relator of R.
- The edges of Γ are between translators and ordinary vertices.

- For each internal translator t, the edges at t are consecutive vertices of one or two parts of Γ, representing subwords u and v of relators of R. And there exists a word w such that $w = \delta_1 u \delta_2 v$ and $w = e$ in G, where $\delta_{1,2}$ have length at most $2E \log \ell$.

- For each boundary translator b, the edges at b are consecutive vertices of one part of Γ, representing a subword u of some relator of R. For each such b, there exists a word w such that $w = \delta_1 u \delta_2 v$ and $w = e$ in G, where v is a subword of the boundary of D, and where $\delta_{1,2}$ have length at most $2E \log \ell$.

- For each commutation translator c, the edges at c are double edges to successive vertices of one part of Γ, representing a subword u of some relator of R. And there exists a word w such that $w = \delta_1 u \delta_2 u^{-1}$ and $w = e$ in G, where $\delta_{1,2}$ have length at most $\varepsilon_2 |u|$.

- There are no double edges except those at commutation translators.

- There are at most $3K/\varepsilon_2$ translators.

- The total number of edges of Γ is at least $N\ell(1 - \varepsilon_1)$ where N is the number of faces of the diagram D bearing a relator of R.

The numbers K and $\varepsilon_1, \varepsilon_2$ are arbitrary. The number E depends on G and K but not on ℓ.

Axioms 2, 3 and 4 are carefully designed to control the probability that, respectively, a boundary translator, internal translator, and commutation translator can be filled.

Note that this graph depends only on the coarsening of the Dehn diagram (up to some dividing done for the elimination of doublets; say we add some decoration to the coarsening indicating how this was done).

Keep all these properties (and notations) in mind for the sequel.

6.7 Apparent length

The line of the main argument below is to fulfill the daDd by filling the translators one by one.

As the same subword of a relator can be joined to a large number of different translators (if the relator appears several times in the diagram), during the construction, at some steps it may happen that one half of a given translator is filled, whereas another part is not. The solution is to remember in one way or another, for each half-filled translator, what is the probability that, given its already-filled side, the word on the other side will fulfill the translator. This leads to the notion of apparent length, which we define now.

Say we are given an element x of the group, of norm $\|x\|$. We try to know if this is a subword of one of our random words under the probability measure μ_ℓ, and to determine the length of this subword.

Given Axiom 2, a good guess for the length of the subword would be $\|x\|/\kappa_2$, with the probability of a longer subword decreasing exponentially.

Given Axiom 3, a good method would be to take another subword y of length $|y|$ at random under μ_ℓ, and test (taking $u = v = e$ in Axiom 3) the probability that $xy = 1$. If x were a subword under μ_ℓ, this probability would be roughly $(2m)^{-\beta(|x|+|y|)}$, hence an evaluation $-\frac{1}{\beta}\log \Pr(xy = e) - |y|$ for the hypothetical length of the subword x.

This leads to the notion of apparent length.

We are to apply Axiom 3 to translators, with u and v of size $2E\log\ell$. For fixed$x \in G$, let $L \geqslant 0$ and denote by $p_L(xuyv = e)$ the probability that, if y is a subword of length L under μ_ℓ (in the sense of Definition 6) there exist words u and v of length at most $2E\log\ell$ such that $xuyv = e$.

Definition 34 (Apparent length at a test-length) – *The* apparent length *of x at test-length L is*

$$\mathbb{L}_L(x) = -\frac{1}{\beta}\log p_L(xuyv = e) - L$$

By definition, if we have a rough evaluation of p_L, we get an evaluation of $\mathbb{L}_L$ up to $o(\ell)$ terms.

We are to apply this definition for y a not too small subword. That is, we will have $\varepsilon_3\ell \leqslant |y| \leqslant \ell$ (for some ε_3 to be fixed soon). We will also use the evaluation from Axiom 2.

Definition 35 (Apparent length) – *The* apparent length *of x is*

$$\mathbb{L}(x) = \min\left(\|x\|/\kappa_2, \min_{\varepsilon_3\ell\leqslant L\leqslant\ell}\mathbb{L}_L(x)\right)$$

Our main tool will now be the following

Proposition 36 – *For any subword x under μ_ℓ, we have*

$$\Pr\left(\mathbb{L}(x) \leqslant |x| - L\right) \lesssim (2m)^{-\beta L}$$

uniformly in L.

As usual, in this proposition, the sense of "for any subword under μ_ℓ" is that of Definition 6.

Proof – This is simply a rewriting of Axioms 2 and 3, combined to the observation that the choice of the test-length and of the small words u and v (which are of length $O(\log\ell)$) only introduces a polynomial factor in ℓ. □

In our proof, we will also need the fact that the number of possible apparent lengths for subwords under μ_ℓ grows subexponentially with ℓ. So we need at least a rough upper bound on the apparent length.

By definition, if x appears with probability p as a subword under μ_ℓ, then by symmetry y will by equal to x^{-1} with the same probability, and we will have that the probability that $xuyv = e$ is at least p^2 (taking $u = v = e$). Thus $\mathbb{L}_{|x|}(x) \leqslant -\frac{2}{\beta}\log p - |x|$. Reversing the equation, this means that for any subword x under μ_ℓ, we have that $\Pr(\mathbb{L}(x) \geqslant L) \leqslant (2m)^{-\beta(L-|x|)/2}$.

In particular, taking L large enough ($L \geqslant 4\ell$ is enough) ensures that in a set of $(2m)^{d\ell}$ randomly chosen elements under μ_ℓ, subwords of length greater than L only occur with probability exponentially small as $\ell \to \infty$. Thus, from now on we suppose that all subwords of words of R have apparent length at most 4ℓ.

In the applications given in this text to plain random words or random geodesic words, apparent length has more properties, especially a very nice behavior under multiplication by a random word. We do not explicitly need these properties, though they are present in the inspiration of our arguments, and thus we do not state them.

6.8 The main argument

Now we proceed to the main step of the proof. We will consider a daDd and evaluate the probability that it is fulfillable. We will see that either the daDd satisfies some isoperimetric inequality, or this probability is very small (exponential in ℓ). It will then be enough to sum on all daDd's with at most K faces to prove Proposition 30.

In our graph Γ, the ordinary vertices represent letters of random relators. Say Γ has $n\ell$ ordinary vertices, that is, the faces of D'' bear n different relators of R.

We will use the term *letter* to denote one of these vertices. Enumerate letters in the ovious way from 1 to $n\ell$, beginning with the first letter of the first relator. So, a letter is a number between 1 and $n\ell$ indicating a position in some relator. Relators are random words on elements of the generating set S of G, so if i is a letter let f_i be the corresponding element of S.

Since the relators are chosen at random, the f_i's are random variables.

As in the case of random quotients of the free group, the idea is to construct the graph Γ step by step, and evaluate the probability that at each step, the conditions imposed by the graph are satisfied by the random set R of relators. We will construct the graph by groups of successive letters joined to the same translators, and use the notion of apparent length (see Definition 34) to keep track of the probabilities involved at each step.

For a letter i, write $i \in t$ if i is joined to translator t. For $1 \leqslant a \leqslant n$, write $i \in a$ to mean that letter i belongs to the a-th part of the graph. So r_a is the product of the f_i's for $i \in a$.

Consider an internal translator t. There is a word w associated to it, which writes $w = u\delta_1 v\delta_2$ where $\delta_{1,2}$ are short and u and v are subwords of the random relators. The subwords u and v are products of letters, say $u = f_p \dots f_q$ and $v = f_r \dots f_s$. Reserve these notations $w(t)$, $u(t)$, $v(t)$, $p(t)$, $q(t)$, $r(t)$ and $s(t)$. Give similar definitions for boundary translators and commutation translators.

Call u and v the *sides* of translator t. The translator precisely imposes that there exist short words δ_1, δ_2 such that $u\delta_1 v\delta_2 = e$ in G. We will work on the probabilities of these events.

Some of the translators may have very small sides; yet we are to apply asymptotic relations (such as the definition of cogrowth) which ask for arbitrarily long words. As there are at most $3K/\varepsilon_2$ translators, with at most two sides each, the total length of the sides which are of length less than $\varepsilon_3\ell$ does not exceed $\varepsilon_3\ell.6K/\varepsilon_2$. Setting $\varepsilon_3 = \varepsilon_2^2/6K$ ensures that the total length of these sides is less than $\varepsilon_2\ell$.

Call *zero-sided translator* an internal translator both sides of which have length less than $\varepsilon_3\ell$. Call *two-sided translator* an internal translator having at least one side of length at least $\varepsilon_3\ell$ and its smaller side of length at least ε_3 times the length of its bigger side. Call *one-sided translators* the rest of internal translators.

Throw away all zero-sided translators from the graph Γ. This throws away a total length of at most $\varepsilon_2\ell$, and do not call sides any more the small sides of one-sided translators. Now we have two-sided translators, one-sided translators, commutation translators and boundary translators, all sides of which have length at least $\varepsilon_3^2\ell$. So we can apply the probability evaluations of Axioms 1-4 without trouble.

For a letter i, say that translator t is finished at time i if $i \geqslant s(t)$. Say that two-sided translator t is half-finished at time i if $q(t) \leqslant i < r(t)$.

Add a further decoration to Γ: for each two-sided translator t, specify an integer $L(t)$ between 0 and 4ℓ (remember we can suppose that every subword has apparent length at most 4ℓ). This will represent the apparent length of the half-word $u(t)$ associated to the diagram when it is half-finished. In the same vein, specify an integer $L(b)$ between 0 and 4ℓ for each boundary translator b, which will represent the apparent length of the word $u(b)$ when b is finished. We want to show that the boundary length is big, so we want to show that these apparent lengths of boundary translators are big. What we will show is the following: if the sum of the imposed $L(c)$'s for all boundary translators c is too small, the probability that the diagram is fulfillable is small.

Now say that a random set of relators $r_1, \dots, r_n$ *fulfills* the conditions of Γ up to letter i if for any internal or commutation translator t which is finished at time i, the corresponding word $w(t)$ is trivial in G; and if, for any half-finished two-sided translator t, the apparent length of the half-word $u(t)$ is $L(t)$; and if, for each finished boundary translator c, the apparent length of $u(c)$ is $L(c)$.

(An apparent length is not necessarily an integer; by prescribing the apparent length of $u(t)$, we prescribe only the integer part. As ℓ is big the discrepancy is totally negligible and we will not even write it in what follows.)

For a given relator r, there may be some translators having a side made of an initial and final piece of r, so that the side straddles the first letter of r. As we will fill in letters one by one starting with the first ones, we should treat these kind of translators in a different way. If a translator side is made of an initial piece and a final piece of some relator, it is enough, for the proof to work, to keep track of the apparent length at the intermediate step when only one part of the side is done. As this leads only to heavier notations, we will neglect this problem.

Of course, fulfillability of the daDd implies fulfillability of Γ up to the last letter for some choice of $r_1, \ldots, r_n \in R$ and for some choice of the $L(t)$'s. (It is not exactly equivalent as we threw away some small proportion ε_1 of the edges.)

The principle of the argument is to look at the evolution of the apparent length of the translators, where the apparent length of a translator at some step is the apparent length of the part of this translator which is filled in at that step. We will show that our axioms imply that when we add a subword of some length, the probability that the increase in apparent length is less than the length of the subword added is exponentially small, such that a simple equation is satisfied:

$$\Delta \mathbb{L} \geqslant |.| + \frac{\Delta \log \mathrm{Pr}}{\beta}$$

(where Δ denotes the difference between before and after filling the subword). This will be the motto of all our forthcoming arguments.

But at the end of the process, the word read on any internal translator is e, which is of apparent length 0, so that the only contribution to the total apparent length is that of the boundary translators, which we therefore get an evaluation of.

Now for a rigorous exposition.

Let P_i be the probability that Γ is fulfilled up to letter i by *some fixed choice* of $r_1, \ldots, r_n \in R$. Note that since all relators of R have the same law μ_ℓ, the quantity P_i does not depend on the choice of relators.

Let $1 \leqslant a \leqslant n$ (recall n is the number of parts of the graph, or the number of different relators of R appearing in the diagram). Let i_0 be the first letter of a, and i_f the last one.

Let P^a be the probability that *there exists* a choice of relators $r_1, \ldots, r_a$ in R fulfilling the conditions of Γ up to letter i_f (the last letter of a). As there are by definition $(2m)^{d\ell}$ choices for each relator, we have

$$\mathrm{P}^a/\mathrm{P}^{a-1} \leqslant (2m)^{d\ell}\mathrm{P}_{i_f}/\mathrm{P}_{i_0-1}$$

which expresses the fact that when we have fulfilled up to part $a-1$, to fulfill up to part a is to choose the a-th relator in R and to see if the letters $f_{i_0}, \ldots, f_{i_f}$ of this relator fulfill the conditions imposed on the a-th part of the graph by the translators.

Let A_a be the sum of all $L(t)$'s for each two-sided translator t which is half-finished at time i_f, plus the sum of all $L(b)$'s for each boundary translator b which is finished at time i_f.

We will study $A_a - A_{a-1}$. The difference is due to internal translators which are half-finished at time i_0 and are finished at time i_f, to internal translators which are not begun at time i_0 and are half-finished at time i_f, and to boundary translators not begun at time i_0 but finished at time i_f: that is, all internal or boundary translators joined to a letter between i_0 and i_f.

First, consider a two-sided translator t which is not begun at time i_0 and half-finished at time i_f. Let $\Delta_t A_a$ be the contribution of this translator to $A_a - A_{a-1}$, we have $\Delta_t A_a = L(t)$ by definition. Taking notations as above, we have an equality $e = u\delta_1 v\delta_2$ in G. By assumption, to fulfill the conditions imposed by Γ we must have $\mathbb{L}(u) = L(t)$. The word u is a subword of the part a of Γ at play. But Proposition 36 tells us that, conditional to whatever happened up to the choice of u, the probability that $\mathbb{L}(u) = L(t)$ is roughly less than $(2m)^{-\beta(|u|-L(t))}$. Thus, taking notations as above, with p the first letter of u and q the last one, we have

$$\mathrm{P}_q/\mathrm{P}_{p-1} \lesssim (2m)^{-\beta(|u|-L(t))}$$

or, taking the log and decomposing u into letters:

$$\Delta_t A_a \geqslant \sum_{i\in t, i\in a} 1 + \frac{\log_{2m}\mathrm{P}_i - \log_{2m}\mathrm{P}_{i-1}}{\beta} + o(\ell)$$

where 1 stands for the length of one letter (!). Note that a rough evaluation of the probabilities gives an evaluation up to $o(\ell)$ of the apparent lengths.

This is the rigorous form of our motto above.

Second, consider an internal translator t which is half-finished at time i_0 and finished at time i_f. Let $\Delta_t A_a$ be the contribution of this translator to $A_a - A_{a-1}$, we have $\Delta_t A = -L(t)$. Taking notations as above, we have an equality $e = u\delta_1 v\delta_2$ in G. By assumption, we have $\mathbb{L}(u) = L(t)$. But the very definition of apparent length (Definition 34) tells us that given u, whatever happened before the choice of v, the probability that there exist such $\delta_{1,2}$ such that $e = u\delta_1 v\delta_2$ is at most $(2m)^{-\beta(\mathbb{L}(u)+|v|)}$. Thus

$$\mathrm{P}_s/\mathrm{P}_{r-1} \lesssim (2m)^{-\beta(L(t)+|v|)}$$

where r and s are the first and last letter making up v. Or, taking the log and decomposing v into letters:

$$\Delta_t A_a \geqslant \sum_{i\in t, i\in a} 1 + \frac{\log_{2m} \mathrm{P}_i - \log_{2m} \mathrm{P}_{i-1}}{\beta} + o(\ell)$$

which is exactly the same as above.

Third, consider an internal translator t which is not begun at time i_0 and finished at time i_f, that is, t is joined to two subwords of the part a of the graph at play. As we removed doublets, the subwords u and v are disjoint, and thus we can work in two times and apply the two cases above, with first t going from not begun state to half-finished state, then to finished state. The contribution of t to $A_a - A_{a-1}$ is 0, and summing the two cases above we get

$$\Delta_t A_a = 0 \geqslant \sum_{i\in t} 1 + \frac{\log_{2m} \mathrm{P}_i - \log_{2m} \mathrm{P}_{i-1}}{\beta} + o(\ell)$$

which is exactly the same as above.

Fourth, consider a commutation translator t which is not begun at time i_0 and is finished at time i_f. Write as above that $e = \delta_1 u \delta_2 u^{-1}$ in G, with δ_1 and δ_2 of length at most $\varepsilon_2 |u|$. By Axiom 4, whatever happened before the choice of u, this event has probability roughly less than $(2m)^{\gamma\varepsilon_2|u|-\beta|u|}$ where γ is some constant. Take $\varepsilon_4 = \gamma\varepsilon_2/\beta$, and as usual denote by p and q the first and last letters making up u. We have shown that

$$\mathrm{P}_q/\mathrm{P}_{p-1} \lesssim (2m)^{-\beta|u|(1-\varepsilon_4)}$$

Take the log, multiply everything by two (since each letter joined to the commutation diagram t is joined to it by a double edge), so that

$$\Delta_t A_a = 0 \geqslant \sum_{i\in t} 2(1-\varepsilon_4) + 2\frac{\log_{2m} \mathrm{P}_i - \log_{2m} \mathrm{P}_{i-1}}{\beta} + o(\ell)$$

Fifth, consider a one-sided translator not begun at time i_0 and finished at time i_f. We have an equality $e = u\delta_1 v\delta_2$ in G, where $\delta_{1,2}$ have length $O(\log \ell)$ and $|v| \leqslant \varepsilon_3 |u|$ (by definition of a one-sided translator), so that $\|u\| \leqslant \varepsilon_3 |u| + O(\log \ell)$. But by Axiom 2, this has probability roughly less than $(2m)^{-\beta|u|(1-\varepsilon_3/\kappa_2)}$, so once again, setting $\varepsilon_5 = \varepsilon_3/\kappa_2$:

$$\Delta_t A_a = 0 \geqslant \sum_{i\in t, i\in a} (1-\varepsilon_5) + \frac{\log_{2m} \mathrm{P}_i - \log_{2m} \mathrm{P}_{i-1}}{\beta} + o(\ell)$$

Sixth and last (guess what and do it yourself), consider a boundary commutator t that is not begun at time i_0 and is finished at time i_f. Its situation is identical to that of an internal translator half-finished at time i_f (first case above), and we get

$$\Delta_t A_a = L(t) \geqslant \sum_{i\in t, i\in a} 1 + \frac{\log_{2m} \mathrm{P}_i - \log_{2m} \mathrm{P}_{i-1}}{\beta} + o(\ell)$$

We are now ready to conclude. Sum all the above inequalities for all translators joined to part a:

$$\begin{aligned} A_a - A_{a-1} &= \sum_{t \text{ translator joined to } a} \Delta_t A_a \\ &\geqslant \sum_{\substack{t \text{ non-commutation translator} \\ i\in t, i\in a}} (1-\varepsilon_5) + \frac{\log_{2m} \mathrm{P}_i - \log_{2m} \mathrm{P}_{i-1}}{\beta} \\ &\quad + \sum_{\substack{t \text{ commutation translator} \\ i\in t, i\in a}} 2(1-\varepsilon_4) + 2\frac{\log_{2m} \mathrm{P}_i - \log_{2m} \mathrm{P}_{i-1}}{\beta} \\ &\quad + o(\ell) \end{aligned}$$

Let m_a be the number of times the a-th relator appears in the Dehn diagram. The way we constructed the graph, any vertex representing a letter of the a-th relator is joined to m_a translators (except for a proportion at most $\varepsilon_1 + \varepsilon_2$ we removed afterwards), counting commutation translators twice. Thus, in the sum above, each of the ℓ letters of a appears exactly m_a times, and so

$$A_a - A_{a-1} \geqslant m_a \left(\ell(1-\varepsilon_4-\varepsilon_5-\varepsilon_1-\varepsilon_2) + \frac{\log_{2m} \mathrm{P}_{i_f} - \log_{2m} \mathrm{P}_{i_0-1}}{\beta} \right) + o(\ell)$$

(Because of the removal of a proportion at most $\varepsilon_1 + \varepsilon_2$ of the letters, some terms $\log_{2m} \mathrm{P}_{i_f} - \log_{2m} \mathrm{P}_{i_0-1}$ are missing in the sum; but as for any i, we have $\mathrm{P}_i \leqslant \mathrm{P}_{i-1}$, the difference of log-probabilities $\log_{2m} \mathrm{P}_i - \log_{2m} \mathrm{P}_{i-1}$ is non-positive, and the inequality is true *a fortiori* when we add these missing terms.)

Note that there is nothing bad hidden in the summation of the $o(\ell)$ terms, since the number of terms in the sum is controlled by the combinatorics of the diagram (i.e. by K), and depends in no way on ℓ.

Recall we saw above that

$$\mathrm{P}^a/\mathrm{P}^{a-1} \leqslant (2m)^{d\ell} \mathrm{P}_{i_f}/\mathrm{P}_{i_0-1}$$

where the $(2m)^{d\ell}$ factor accounts for the choice of the relator r_a in R.

Set $d_a = \log_{2m} \mathrm{P}^a$ (compare the case of random quotients of F_m). Beware the d_a's are negative.

Setting $\varepsilon_6 = \varepsilon_4 + \varepsilon_5 + \varepsilon_1 + \varepsilon_2 + o(\ell)/\ell$ we get

$$A_a - A_{a-1} \geqslant m_a \left(\ell(1-\varepsilon_6) + \frac{d_a - d_{a-1} - d\ell}{\beta} \right)$$

Compare this to the equation linking dimension and number of edges on page 16 (and recall that here A_a is not the number of edges but the apparent length, which varies the opposite way, and that we want it to be big).

Summing the inequalities above for a from 1 to n gives

$$\begin{aligned} A_n &\geqslant \ell(1-\varepsilon_6)\sum m_a - \frac{d\ell}{\beta}\sum m_a + \frac{1}{\beta}\sum m_a(d_a - d_{a-1}) \\ &= |D''|\,\ell\left(1-\varepsilon_6-\frac{d}{\beta}\right) + \frac{1}{\beta}\sum d_a(m_a - m_{a+1}) \end{aligned}$$

But at the end of the process, all translators are finished, so A_n is simply the sum of the apparent lengths of all boundary translators, that is $A_n = \sum_b L(b)$.

Now recall that a boundary translator b means the existence of an equality $e = \delta_1 u \delta_2 v$ in G, with by assumption $\mathbb{L}(u) = L(b)$, the δ's of length at most $2E\log\ell$, and v lying on the boundary of the diagram. By the definition of apparent length (Definition 35 which takes Axiom 2 into account), we have $\|u\| \geqslant \kappa_2\mathbb{L}(u) = \kappa_2 L(b)$, thus $\|v\| \geqslant \|u\| - \|\delta_1\| - \|\delta_2\| \geqslant \kappa_2 L(b) + o(\ell)$. As v lies on the boundary of D this implies

$$|\partial D| \geqslant \kappa_2 A_n + o(\ell)$$

(Recall we can sum the $o(\ell)$'s harmlessly since the number of translators is bounded by some function of K.)

So, setting $\kappa_3 = \kappa_2/\beta - o(\ell)/\ell$ and using the lower bound for A_n above we get

$$|\partial D| \geqslant |D''|\,\ell\,(\beta(1-\varepsilon_6) - d)\,\kappa_3 + \kappa_3 \sum d_a(m_a - m_{a+1})$$

Now choose ε_1, ε_2 and ε_3 small enough (depending on K, β, κ_2 and d but not on ℓ), in such a manner that $\beta(1-\varepsilon_6) - d \geqslant (\beta - d)/2$. (For example, take $\varepsilon_6 \leqslant \frac{1-d/\beta}{2}$, which is possible since $d < \beta$). The equation above rewrites

$$|\partial D| \geqslant |D''|\,\ell\,(\beta - d)\,\kappa_3/2 + \kappa_3 \sum d_a(m_a - m_{a+1})$$

We are free to choose the order of the enumeration of the parts of the graph. In particular, we can suppose that the m_a's are non-increasing.

As $\sum m_a = |D''|$, we have $\sum d_a(m_a - m_{a+1}) \geqslant |D''| \inf d_a$ (recall the d_a's are non-positive). Thus

$$|\partial D| \geqslant \frac{\kappa_3}{2}\,|D''|\,\ell\,(\beta - d + 2\inf d_a/\ell)$$

By definition, the probability that the daDd is fulfillable is less than $(2m)^{d_a}$ for all a. This probability is then less than $(2m)^{\inf d_a}$.

First suppose that $\inf d_a \geqslant -\ell(\beta - d)/4$. Then we have the isoperimetric inequality

$$|\partial D| \geqslant \frac{\kappa_3}{4} \ell \, |D''| \, (\beta - d)$$

To put it in the exact form of Proposition 30, recall that $|D'| \leqslant 4\,|D''|\,\ell$ by assumption, and then write $|\partial D| \geqslant \frac{\kappa_3}{8}\ell(\beta - d)\,|D''| + \alpha'\,|D'|$ where $\alpha' = (\beta - d)\kappa_3/32$.

Or, second, suppose $\inf d_a < -\ell(\beta - d)/4$ and this means that the probability that the daDd is fulfillable is less than $(2m)^{-\ell(\beta-d)/4}$, which decreases exponentially in ℓ.

In order to prove Proposition 30 we have not only to evaluate this probability for one daDd but for all daDd's having at most K faces. Note that a daDd is given by a planar graph with at most K faces and lots of decoration on it. The decoration consists of numbers between 1 and K on each face, several lengths between 1 and ℓ on each edge (to define translators and to keep track of the elimination of doublets), and a length between 1 and 4ℓ on some translators (to assign apparent lengths); the number of lengths to be put is controlled by K and ε_2 and does not depend on ℓ. The number of choices for the decoration is thus polynomial in ℓ. Multiplying by the (finite!) number of planar graphs having at most K faces proves that the probability that there exists a daDd violating the isoperimetric inequality decreases exponentially with ℓ.

This proves Proposition 30, hence hyperbolicity of the random quotient when $d < \beta$.

6.9 Non-elementarity of the quotient

We now prove that if $d < \beta$, the quotient is infinite and not quasi-isometric to $\mathbb{Z}$.

6.9.1 Infiniteness

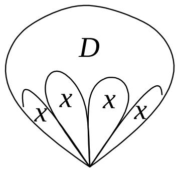
Let $d < \beta$. We will show that the probability that the random quotient is finite decreases exponentially as $\ell \to \infty$.

We know from hyperbolicity of the quotient (Proposition 31) that the probability that there exists a Dehn diagram of the quotient whose part made of old relators is reduced and which is strongly reduced with respect to G, violating some isoperimetric inequality, is exponentially close to 0.

Imagine that $G/\langle R \rangle$ is finite. Then any element of the quotient is a torsion element. Let x be an element of the quotient, this means that there exists a Dehn diagram D bordered by x^n for some n. This diagram can be reduced, and it can also be strongly reduced by Proposition 29. So we can suppose the diagram is strongly reduced.

Now take for x a random word picked under μ_ℓ. Instead of applying the previous section's results to the random quotient of G by R, consider the random quotient of G by $R \cup \{x\}$. Since x is taken at random, $R \cup \{x\}$ is a random set of words, whose density is only slightly bigger than d; this density is $d' = \frac{1}{\ell}\log_{2m}\left((2m)^{d\ell}+1\right)$ which, if ℓ is large enough, is smaller than β if d is.

Now, if $G/\langle R\rangle$ is finite then x is of torsion. Set $N = |R| = (2m)^{d\ell}$. Consider the following family of diagrams. Let D be any abstract Dehn diagram of $G/\langle R\rangle$ of boundary length $n\ell$ for some n. Define the spherical diagram E by gluing n faces of boundary size ℓ on the boundary of D along their border, and associate to each of the new faces the relator number $N+1$, so that D is an abstract Dehn diagram with respect to $R \cup \{x\}$. If $G/\langle R\rangle$ is finite then x is of torsion, thus at least one of the diagrams E is fulfillable with respect to $R \cup \{x\}$.



The resulting diagram is strongly reduced as we can suppose that D is and since we only added faces bearing number $N+1$ with the same orientation.

So there exists a strongly reduced Dehn diagram of $G/\langle R \cup \{x\}\rangle$ with boundary 0.

But we know that, in the random quotient $G/\langle R \cup \{x\}\rangle$ at density d', the existence of such a diagram is of probability exponentially close to 0 as ℓ tends to infinity. This ends the proof.

6.9.2 Non-quasi$\mathbb{Z}$ness

We show here that the random quotients for $d < \beta$ are not quasi-isometric to $\mathbb{Z}$. Of course, we suppose $\beta > 0$, which amounts, in the case we consider (plain, or reduced, or geodesic words), to G itself not being quasi-isometric to $\mathbb{Z}$.

We will reason in a similar manner as above to prove infiniteness. We will consider a random quotient by a set R of words at density d, and we will add to R two random words x and y picked under μ_ℓ, thus obtaining a new random set of words at a density $d' > d$. As ℓ is big, d' is only slighlty above d, and if ℓ is big enough we still have $d' < \beta$.

Say (from Proposition 31) that any strongly reduced diagram D of the group $G/\langle R'\rangle$ satisfies an isoperimetric inequality $|\partial D| \geqslant \alpha\ell\,|D''|$ for some positive α, notations as above.

Suppose that $G/\langle R\rangle$ is quasi-isometric to $\mathbb{Z}$.

The two random elements x and y are either torsion elements or each of them generates a subgroup of finite index. The case of torsion is handled exactly as above in the proof of infiniteness.

Thus, suppose x is not a torsion element. Let h be the index of the subgroup it generates. Of course h depends on x.

For any $n \in \mathbb{Z}$, we can find a p such that $y^n = x^p u$ in $G/\langle R \rangle$, where u is of length at most h. This equality defines a Dehn diagram of $G/\langle R \rangle$. We can take its strong reduction.

Now glue n faces containing y and p faces containing x to the boundary of this diagram. This defines a Dehn diagram of $G/\langle R' \rangle$, and it is still strongly reduced as relators x and y did not appear in the rest of the diagram. The boundary of this diagram is u.

But n can be taken arbitrarily large, so we can take $n > |u|/\alpha$. THen the diagram has at least n faces and boundary length $|u|$, which contradicts our isoperimetric inequality $|\partial D| \geqslant \alpha\ell\,|D''|$.

Of course, u, n and p depend on the random words x and y. But consider the following family of diagrams: for each $h \in \mathbb{N}$, each $p \in \mathbb{N}$ and each $n \in \mathbb{N}$ such that $n > h/\alpha$, consider all abstract Dehn diagrams of length $h + n\ell + p\ell$, with the numbers on the faces between 1 and $N = |R|$. Consider the diagrams obtained from these by the following process : glue p faces of size ℓ bearing number $N+1$ on the boundary, and n faces of size ℓ bearing number $N+2$.

The reasoning above shows that if $G/\langle R \rangle$ is quasi-isometric to $\mathbb{Z}$, then at least one of these abstract Dehn diagrams is fulfillable by a strongly reduced Dehn diagram on the relators of R'. But all these diagrams violate the isoperimetric inequality, hence the conclusion.

Alternate proof. We give an alternate proof as it uses an interesting property of the quotients. It works only in the case of a random quotient by uniformly chosen plain words.

Proposition 37 – *If $d > 0$, then the abelianized of a random quotient of any group by uniformly chosen plain random words is (with probability arbitrarily close to 1 as $\ell \to \infty$) either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$.*

(As usual, we find $\mathbb{Z}/2\mathbb{Z}$ when ℓ is even and there are no relations of odd length in the presentation of G.)

Of course this is not necessarily true if $d = 0$, since in this case the number of relations added does not tend to infinity.

Proof –

We want to show that a random quotient in density $d > 0$ of the free abelian group $\mathbb{Z}^m$ is trivial or $\mathbb{Z}/2\mathbb{Z}$.

Take a random word of length ℓ on $a_1^{\pm 1}, \ldots a_m^{\pm 1}$. By the central limit theorem (or by an explicit computation on the multinomial distribution), the number of times generator a_i appears is roughly $\ell/2m$ up to $\pm\sqrt{\ell}$.

For the sake of simplicity, say that ℓ is a multiple of $2m$. The probability that a random word w is such that all relators a_i and a_j^{-1} appear exactly $\ell/2m$ times in w is equivalent to (by the central limit theorem with $2m-1$ degrees of freedom or by a direct computation using Stirling's formula)

$$\frac{\sqrt{2m}}{(\pi\ell/m)^{(2m-1)/2}}$$

This is equivalent as well to the probability that all a_i and a_j^{-1} appear exactly $\ell/2m$ times, except for some a_{i_0} appearing $1+\ell/2m$ times and some a_{j_0} appearing $\ell/2m-1$ times, this deviation being negligible.

This probability decreases polynomially in ℓ. But we choose an exponential number of random words, namely $(2m)^{d\ell}$. So if $d>0$, with very high probability we will choose a word w in which all a_i and a_j^{-1} appear exactly $\ell/2m$ times, except for some a_{i_0} appearing $1+\ell/2m$ times and some a_{j_0} appearing $\ell/2m-1$ times.

But $w=e$ in the quotient, and $w=e$ in an abelian group is equivalent to $a_{i_0}a_{j_0}^{-1}=e$ since all other relators appear exactly the same number of times with exponent 1 or -1 and thus vanish.

As this occurs arbitrarily many times, this will happen for all couples of i,j. So these relators satisfy $a_i=a_j^{\pm1}$ in the quotient for all i,j. In particular, all relators are equal and moreover we have $a_i=a_i^{-1}$.

Thus the abelianized is either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$. $\square$

Corollary 38 – *A random quotient of a hyperbolic group by plain random words for $d<\beta$ is not quasi-isometric to $\mathbb{Z}$.*

Proof – First take $d>0$. It is well-known (cf. [SW], Theorem 5.12, p. 178) that a group which is quasi-isometric to $\mathbb{Z}$ has either $\mathbb{Z}$ or the infinite diedral group D_∞ as a quotient.

If $\mathbb{Z}$ is a quotient of the group, then its abelianized is at least $\mathbb{Z}$, which contradicts the previous proposition. If D_∞ is a quotient, note that the abelianized of D_∞ is $D_2=\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/2\mathbb{Z}$, which is still incompatible with the previous proposition. So we are done if $d>0$.

Now if $d=0$, note that a random quotient with $d>0$ is a quotient of a random quotient with $d=0$ (isolate the first relators). If the random group at $d=0$ were quasi-isometric to $\mathbb{Z}$, then all of its quotients would be either finite or quasi-isometric to $\mathbb{Z}$, which is not the case. (Note that here we use hyperbolicity of G to be allowed to apply our main theorem, implying that random quotients are non-trivial for some $d>0$. It may be that random quotients at $d=0$ of some groups are quasi-isometric to $\mathbb{Z}$.) $\square$

This ends the proof of Theorem 9.

A Appendix: The Cartan-Hadamard-Gromov-Papasoglu theorem

The Cartan-Hadamard-Gromov-Papasoglu theorem allows to go from a local isoperimetric inequality (concerning small figures in a given space) to isoperimetry at large scale. It lies at the heart of our argument: to ensure hyperbolicity of a group, it is enough to check the isoperimetric inequality for a finite number of diagrams. This finite number depends, of course, of the quality of the isoperimetric inequality we get on these small diagrams. In particular, there is an algorithm to detect hyperbolicity of a given group (see [Pap]).

Let us state the form of the theorem we will use.

Let X be a simplicial complex of dimension 2 (all faces are triangles). A *circle drawn in* X is a sequence of consecutive edges such that the endpoint of the last edge is the starting point of the first one. A *Disk dran in* X is a simplicial map from a triangulated disk to X.

The *area* A_{tr} of a disk drawn in X is its number of triangles. The *length* L_{tr} of a circle drawn in X is its number of edges. (Both with multiplicity.) This is, X is considered being made of equilateral triangles of side 1 and area 1.

The *area* of a drawn circle will be the smallest area of a drawn disk with this circle as boundary, or ∞ if no such disk exists. The *length* of a drawn disk will be the length of its boundary.

Theorem 39 (P. Papasoglu, cf. [Pap]) – *Let X be a simplicial complex of dimension 2, simply connected. Suppose that for some integer $K > 0$, any circle S drawn in X whose area lies between $K^2/2$ and $240K^2$ satisfies*

$$L_{\mathrm{tr}}(S)^2 \geqslant 2 \cdot 10^4 \, A_{\mathrm{tr}}(S)$$

Then any circle S drawn in X with $A(S) \geqslant K^2$ satisfies

$$L_{\mathrm{tr}}(S) \geqslant A_{\mathrm{tr}}(S)/K$$

This theorem is a particular case of a more general theorem stated by Gromov in [Gro1], section 6.8.F, for a length space. Think of a manifold. At very small scales, every curve in it satisfies the same quadratic isoperimetric inequality as in the Euclidean space, with constant 4π. The theorem means that if, at a slightly larger scale, the constant in this quadratic isoperimetric inequality becomes better ($2\cdot 10^4$ instead of 4π), then isoperimetry propagates to large scales, and at these large scales the isoperimetric inequality even becomes linear. This is analogous to the fact that a control on the curvature of a manifold (which is a local invariant) allows to deduce global hyperbolicity properties. This was termed by Gromov *hyperbolic Cartan-Hadamard theorem* or *local-to-global principle for hyperbolic spaces.*

The proof of Papasoglu is based on considering the smallest diagram violating the inequality to prove, and, by some surgery involving only cutting it in various ways, to exhibit a smaller diagram violating the assumptions. As this process only requires to consider subdiagrams of a given diagram, he prooves a somewhat stronger theorem.

Theorem 40 (P. Papasoglu, cf. [Pap]) – *Let X be a simplicial complex of dimension 2, simply connected. Let P be a property of disks in X such that any subdisk of a disk having P also has P.*

Suppose that for some integer $K > 0$, any disk D drawn in X having P, whose area lies between $K^2/2$ and $240K^2$ satisfies

$$L_{\mathrm{tr}}(D)^2 \geqslant 2 \cdot 10^4 \, A_{\mathrm{tr}}(D)$$

Then any disk D drawn in X, having P, with $A(D) \geqslant K^2$, satisfies

$$L_{\mathrm{tr}}(D) \geqslant A_{\mathrm{tr}}(D)/K$$

In the previous version, property P was "having the minimal area for a given boundary", hence the change from circles to disks.

We need to extend these theorems to complexes in which not all the faces are triangular.

Let X be a complex of dimension 2. Let f be a face of X.

The *combinatorial length* L_c of f is defined as the number of edges of its boundary. The *combinatorial area* A_c of f is defined as $L_c(f)^2$.

Let D be a disk drawn in X. The *combinatorial length* L_c of D is the length of its boundary. The *combinatorial area* A_c of D is the sum of the combinatorial areas of its faces.

Proposition 41 – *Let X be a complex of dimension 2, simply connected. Suppose that a face of X has at most ℓ edges. Let P be a property of disks in X such that any subdisk of a disk having P also has P.*

Suppose that for some integer $K \geqslant 10^{10}\ell$, any disk D drawn in X having P, whose area lies between $K^2/4$ and $480K^2$ satisfies

$$L_c(D)^2 \geqslant 2 \cdot 10^{14} \, A_c(D)$$

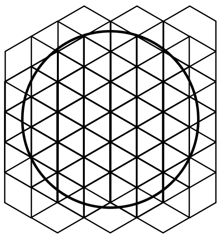
Then any disk D drawn in X, having P, with $A(D) \geqslant K^2$, satisfies

$$L_c(D) \geqslant A_c(D)/10^4 K$$

Proof of the proposition – Of course, we will show this proposition by triangulating X and applying Papasoglu's theorem.

The naive triangulation (cut a n-gon into $n-2$ triangles) does not work since all triangles do not have the same size.

Triangulate all faces of X in the following way: consider a face of X with n sides as a regular n-gon of perimeter n in the Euclidean plane. Consider a triangulation of the plane by equilateral triangles of side 1. (The polygon is drawn here with large n, so that it looks like a circle.)



This is not exactly a triangulation, but with a little work near the boundary, we can ensure that the polygon is triangulated in such a way that all triangles have sides between, say, $1/10$ and 10 and area between $1/10$ and 10, so that the distortion between the triangle metric and the Euclidian metric is a factor at most 10. Note that the number of triangles lies between $n^2/100$ and $100n^2$, as the (Euclidian) area of our n-gon is roughly $n^2/4\pi$.

Let Y be the simplicial complex resulting from X by this triangulation.

Let L_{tr} and A_{tr} be the length and area in Y assigning length 1 to each edge and area 1 to each triangle. Let L_c and A_c be the length and area in X defined above; in Y they can be used for disks coming from X.

Let L and A be the Euclidean length and area in Y, that is, each face of X with n edges is a regular n-gon, and the triangles are given their length and area coming from the triangulation above in the Euclidean plane.

The discrepancy between L_{tr}, L and L_c, and between A_{tr}, A and A_c, is at most a factor 100.

We proceed as follows: We will show that a disk in Y with property P, whose area $A_{\mathrm{tr}}(B)$ lies between $K^2/2$ and $240K^2$, satisfies $L_{\mathrm{tr}}(B)^2 \geqslant 2\cdot 10^4 A_{\mathrm{tr}}(B)$. Then, by the above theorem, any disk B of area $A_{\mathrm{tr}}(B) \geqslant K^2$ will satisfy $L_{\mathrm{tr}}(B) \geqslant A_{\mathrm{tr}}(B)/K$, thus $L_c(B) \geqslant A_c(B)/10^4K$ and we will be done.

Let B be a disk in Y with property P, whose area $A_{\mathrm{tr}}(B)$ lies between $K^2/2$ and $240K^2$. We want to show that it satisfies $L_{\mathrm{tr}}(B)^2 \geqslant 2\cdot 10^4 A_{\mathrm{tr}}(B)$.

There are two kinds of disks drawn in Y: those who come from a disk drawn in X, and those which there exists faces of X that are only partially contained in.

For the first kind we are done: by assumption, we have $L_c(B)^2 \geqslant 2\cdot 10^{14} A_c(B)$, which implies $L_{\mathrm{tr}}(B)^2 \geqslant 2\cdot 10^4 A_{\mathrm{tr}}(B)$.

So we want to reduce the problem to this kind of disks.

We will need the following isoperimetric lemmas:

Lemma 42 – *Let C be a regular closed curve in a Euclidean disk D. Suppose that C encloses a surface of area at most half the area of D. Then the* *length of the intersection of C with the boundary of D is at most 32 times the length of the intersection of C with the interior of D.*

(One would expect an optimal constant $\pi/2$ with optimal C enclosing a half disk.)

This lemma is shown in [Gro3], 6.23. The next lemma is a formal consequence thereof.

Lemma 43 – *Let C be a regular closed curve in a Euclidean disk D. Suppose that C encloses a surface of area at least half the area of D. Then the length of the intersection of C with the interior of D is at least $1/32$ times the length of $\partial D \setminus C$.*

The next lemma is a consequence of the first one and of the usual isoperimetric inequality in the Euclidean plane.

Lemma 44 – *Let C be a regular closed curve in a Euclidean disk D. Suppose that C encloses a surface of area at most half the area of D. Then the square of the length of the intersection of C and the interior of D is at least $1/100$ times the area enclosed by C.*

Now back to our disk B in Y.

Let D be a face of X such that B intersects D.

Suppose that $\partial B \cap D$ is connected (that is, B intersects D only once; otherwise, make the following construction for each of the connected components). Compare the Euclidean area of $B \cap D$ with that of D. If it is more than one half, enlarge B such that it includes all of D.

Follow this process for each face D of X partially intersecting B.

Let B' be the disk in Y obtained after this process. By construction, we have $A(B) \leqslant A(B') \leqslant 2A(B)$. By Lemma 43, we have $L(B') \leqslant 32L(B)$.

Now, for each face D of X intersecting B', either $D \subset B'$ or the area of $D \cap B'$ is at most one half the area of D.

As a first case, suppose that the cumulated area of all such D which are included in B' is at least one half of the area of B'. Define B'' by amputing B' from all faces D of X which are not totally included in B'. By assumption, we have $A(B') \geqslant A(B'') \geqslant A(B')/2$. And it follows from Lemma 42 that $L(B'') \leqslant 32L(B')$.

By construction, the disk B'' is now a disk made of whole faces of X. As $A(B)/2 \leqslant A(B'') \leqslant 2A(B)$, we have $K^2/4 \leqslant A(B'') \leqslant 480K^2$. We can thus apply the isoperimetric assumption: $L(B'')^2 \geqslant 2 \cdot 10^{14} A(B'')$. Since $L(B'') \leqslant 32^2 L(B)$ and $A(B) \leqslant 2A(B'')$, we get that $L(B)^2 \geqslant 2 \cdot 10^{10} A(B)$, hence $L_{\mathrm{tr}}(B) \geqslant 2 \cdot 10^4 A_{\mathrm{tr}}(B)$.

As a second case, imagine that the cumulated area of all such D which are wholly included in B' is less than half the area of B'. Let D_i be the faces of X intersecting B' but not wholly contained in B'. Let $a_i = A(D_i \cap B')$. We have $\sum a_i \geqslant A(B')/2 \geqslant K^2/4$.

Let $m_i = L(\partial B' \cap D_i)$. By Lemma 44, we have $m_i^2 \geqslant a_i/100$.

Since any face of X has at most ℓ edges, we have $A_c(D_i \cap B') \leqslant \ell^2$, so for any i, $a_i \leqslant 100\ell^2$. Group the indices i in packs I so that for each I, we have $100\ell^2 \leqslant \sum_{i \in I} a_i \leqslant 200\ell^2$. There are at least $K^2/800\ell^2$ packs I. Let $M_I = \sum_{i \in I} m_i$.

We have

$$M_I = \sum_{i \in I} m_i \geqslant \sqrt{\sum_{i \in I} m_i^2} \geqslant \sqrt{\sum_{i \in I} a_i/100} \geqslant \ell$$

and

$$L(B')^2 \geqslant \left(\sum_i m_i\right)^2 = \left(\sum_I M_I\right)^2 \geqslant \left(\sum_I \ell\right)^2$$

and as there are at least $K^2/800\ell^2$ packs

$$L(B')^2 \geqslant K^4/10^6\ell^2 \geqslant A(B')K^2/10^9\ell^2$$

as $A(B') \leqslant 480K^2$. Now as $L(B') \leqslant 32L(B)$ and $A(B') \geqslant A(B)$ we have

$$L(B)^2 \geqslant A(B)K^2/10^9\ell^2$$

or

$$L_{\mathrm{tr}}(B)^2 \geqslant A_{\mathrm{tr}}(B)K^2/10^{15}\ell^2$$

and we are done as $K^2 \geqslant 10^{20}\ell^2$.

This ends the proof of the proposition. □

B Appendix: Conjugacy and isoperimetry in hyperbolic groups

We prove here some of the statements needed in the text about conjugacy of words and narrowness of diagrams in hyperbolic groups.

Throughout this appendix, G will denote a hyperbolic discrete group generated by a finite symmetric set S, defined by a finite set of relations R (every discrete hyperbolic group is finitely presented, cf. [S]). Let δ be a hyperbolicity constant w.r.t. S.

A *word* will be a word made of letters in S. The *length* of a word w will be its number of letters (regardless of whether it is equal to a shorter word in the group), denoted by $|w|$.

Equality of words will always be with respect to the group G.

Let C be an isoperimetric constant for G, i.e. a number such that any simply connected minimal Dehn diagram D on G satisfies $|\partial D| \geqslant C\,|D|$. See section 1 for definitions and references about diagrams and isoperimetry.

Let us also suppose that the relations in the presentation R of G have length at most λ.

B.1 Conjugate words in G

The goal of this section is to show that if a word x is known to be a conjugate in G of a short word y, then some cyclic permutation of x is conjugate to y by a short word. If $x = uyu^{-1}$, we will say that x is *conjugate to y by u*, or that *u conjugates x* and y, or that u is a *conjugating word*. We recall the

Definition – *A word w is said to be* fully reduced *if it and all of its cyclic permutations label geodesic words in G.*

The following is well-known (cf. [BH] p. 452).

Proposition 45 – *Let u, v be fully reduced words representing conjugate elements of G. Then*

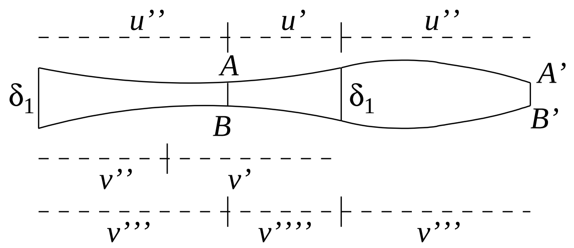
- *either $|u| \leqslant 8\delta + 1$ and $|v| \leqslant 8\delta + 1$*
- *or else there exist cyclic permutations u' and v' of u and v which are conjugate by a word of length at most $2\delta + 1$.*

This immediately extends to:

Proposition 46 – *Let u, v be fully reduced words representing conjugate elements of G. Then*

- *either $|u| \leqslant 8\delta + 1$ and $|v| \leqslant 8\delta + 1$*
- *or else there exist a cyclic permutation v' of v which is conjugate to u by a word of length at most $4\delta + 1$.*

PROOF – Write $u = u'u''$ and $v = v'v''$ such that the cyclic conjugates $u''u'$ and $v''v'$ are conjugate by a word δ_1 of length at most $2\delta + 1$ as in Proposition 45. Construct the quadrilateral $u''u'\delta_1 v'^{-1}v''^{-1}\delta_1^{-1}$. As u and v are fully reduced, the sides $u''u'$ and $v''v'$ are geodesic, and in this δ-hyperbolic quadrilateral any point on one side is 2δ-close to some other side. In particular, any point on the side $u''u'$ is $(2\delta + |\delta_1|)$-close to the side $v''v'$.



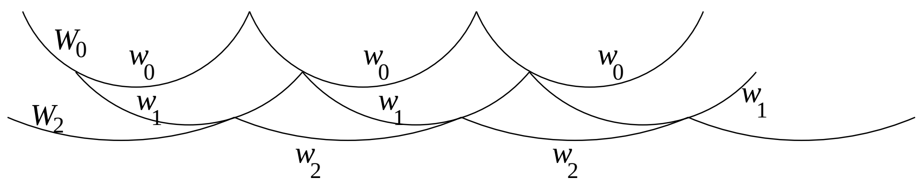
Let A be the endpoint of u''. The point A is $(2\delta + |\delta_1|)$-close to some point B on $v''v'$. Let δ_2 be a path connecting A to B. The point B divides $v''v'$ into two words v''' and v'''', and we have $u = u'u'' = \delta_2 v''''v'''\delta_2^{-1}$ which ends the proof of the proposition. □

We will need the following

PROPOSITION 47 – *Let w be a geodesic word. There exists a fully reduced word z which is conjugate to w by a word of length at most $(|w| - |z|)(\delta + 1/2) + 4\delta$.*

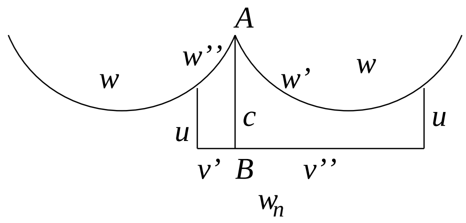
PROOF – Set $w_0 = w$ and construct a sequence w_n of geodesic words by induction. If w_n is fully reduced, stop. If not, then write $w_n = w_n'w_n''$ such that $w_n''w_n'$ is not geodesic. Then set w_{n+1} to a geodesic word equal to $w_n''w_n'$. As length decreases at least by 1 at each step, the process stops after a finite number n of steps and w_n is fully reduced. Note that $n \leqslant |w| - |w_n|$.

In the Cayley graph of the group, define W_i to be the quasi-geodesic $(w_0'w_1' \dots w_{i-1}'w_i^k)_{k\in\mathbb{Z}}$ with w_i' as above:



Consider any of the geodesic triangles made by w_i, w_{i-1}'', w_{i-1}'. As these are δ-hyperbolic, this means that any point of W_i is δ-close to the line W_{i-1}. Thus, any point of W_n is $n\delta$-close to W_0.

Consider the two endpoints of a copy of w_n lying on W_n. These two points are $n\delta$-close to W_0, and since the whole picture is invariant by translation, this means that we can find a word u of length at most $n\delta$ such that u conjugates w_n to some cyclic conjugate $w''w'$ of w. Now construct the hexagon $w''w'uw_n^{-1}u^{-1}$.



Let A be the endpoint of w''. By elementary δ-hyperbolic geometry (approximation by a tripod of the triangle consisting of A and the endpoints of v), the distance of A to the side v is less than $(|w''|+|w'|+2\,|u|-|w_n|)/2+4\delta$. Let B be a point on side w_n realizing this minimal distance. Let $w_n = v'v''$ such that the endpoint of v' is B. Let c be the word defined by AB. Then we have $w'w'' = cv''v'c^{-1}$, so w is conjugate to a cyclic conjugate of w_n by c. Taking $z = v''v'$ ends the proof of the proposition. □

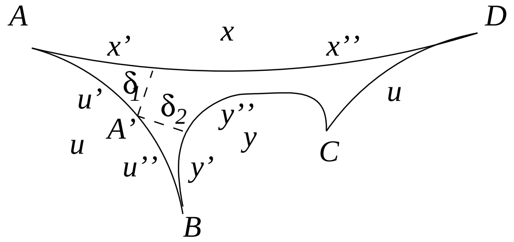
Now, in the spirit of Proposition 45, let $C_c = \max_{x,y} \min\{|u|\,, x = uyu^{-1}\}$ where the range of the maximum is the set of all couples of conjugate words of length at most $8\delta + 1$. As this set is finite we have $C_c < \infty$. Let $C'_c = C_c + 4\delta^2 + 12\delta + 2$.

Proposition 48 – *Let x be a geodesic word and y a conjugate of x of minimal length. Then some cyclic conjugates of x and y are conjugate by a word of length at most C'_c.*

Proof – Let u be a conjugating word of minimal length: $x = uyu^{-1}$. This defines a Dehn diagram $ABCD$ whose sides are labeled by u, y, u^{-1} and x^{-1} in this order.

As x, y and u are geodesic words (by minimality assumption), the 1-skeleton of this diagram embeds in the Cayley graph of the group, and we get a hyperbolic quadrilateral $ABCD$ in which every point on any side is 2δ-close to a point on another side.

As a first case, suppose that every point on the side AB is 2δ-close to either AD or BC.

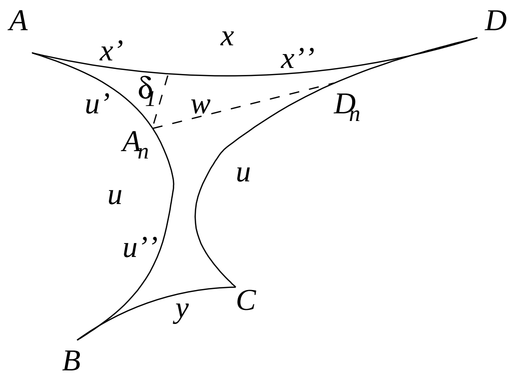


Let A' be the first point on AB which is 2δ-close to BC. Considering the point just before A', we know that A' is $(2\delta + 1)$-close to AD.

Then we can write $x = x'x''$, $u = u'u''$ and $y = y'y''$ such that there exist words δ_1 and δ_2 of length at most $2\delta+1$ such that $u' = x'\delta_1$ and $u'' = \delta_2 y'^{-1}$. Then, we have $x''x' = x'^{-1}xx' = \delta_1 u'^{-1}uyu^{-1}u'\delta_1{}^{-1} = \delta_1 u''yu''^{-1}\delta_1^{-1} = \delta_1\delta_2 y''y'\delta_2^{-1}\delta_1^{-1}$, and the cyclic conjugate $x''x'$ of x is conjugate to $y''y'$ by a word of length at most $4\delta + 2$.

By symmetry the same tricks work if DC is close to DA or to CB.

Second, if this is not the case, let A_n and D_n be the points on AB and DC at distance n away from A and D, respectively. Let n be smallest such that either A_n or D_n is not 2δ-close to AD nor to BC. By symmetry, let us suppose it is A_n rather than D_n. Let w be a geodesic word joining A_n to D_n.



Let u' be the prefix of u joining A to A_n. By definition of n the point A_n is $2\delta + 1$-close to AD. We have $u' = x'\delta_1$ where x' is a prefix of x, and $|\delta_1| \leqslant 2\delta + 1$. Thus $x''x'$ is conjugate to w by a word of length less than $2\delta + 1$.

Now let us work in A_nBCD_n. By definition of A_n, we know there exists a point A' on CD_n such that $A_nA' \leqslant 2\delta$. Now we have $A_nD_n \leqslant 2\delta + A'D_n = 2\delta + D_nC - A'C = 2\delta + A_nB - A'C \leqslant 4\delta + A'B - A'C \leqslant 4\delta + BC$. Thus $|w| \leqslant 4\delta + |y|$.

By our minimality assumption, y is fully reduced. If w is fully reduced as well, then we conclude by Proposition 46. If not, use Proposition 47 to find a fully reduced word z which is conjugate to w by a word of length at most $(|w| - |z|)(\delta + 1/2) + 4\delta$. By our minimality assumption on y, we have that $|z| \geqslant |y|$, hence $|w| - |z| \leqslant |w| - |y| \leqslant 4\delta$. Now z and y are both fully reduced and we conclude by Proposition 46. □

Corollary 49 – *Let x be any word and y be a conjugate of x of minimal length. Then some cyclic conjugates of x and y are conjugate by a word of length at most $\delta \log_2 |x| + C'_c + 1$.*

Proof – This is because in a hyperbolic space, a geodesic joining the ends of any curve of length ℓ stays at distance at most $1 + \log_2 \ell$ from this curve (cf. [BH], p. 400). Take a geodesic word x' equal to x and apply the above proposition, then any cyclic permutation of x' will be conjugate to a cyclic permutation of x by a word of length at most $1 + \log_2 |x|$. □

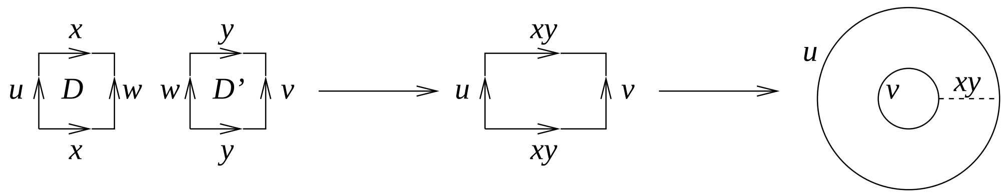
B.2 One-hole diagrams

We now turn to the study of isoperimetry of Dehn diagrams with exactly one hole. Recall that the conjugacy of two words u and v is equivalent to the existence of a one-hole Dehn diagram bordered by u and v.

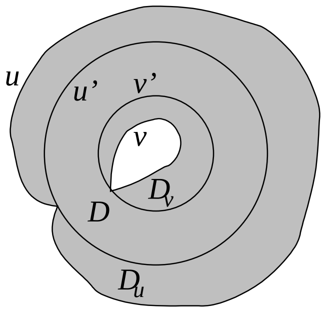
PROPOSITION 50 – *There exists a constant C' such that for any two conjugate words u and v, there exists a one-hole diagram D bordered by u and v, such that $C'|D| \leqslant |u| + |v|$.*

PROOF – Let us first suppose that u and v are geodesic words. Let w be the shortest common conjugate of u and v. By Proposition 48, u and w are conjugate by a word x of length at most $|u|/2 + |w|/2 + C'_c$. Thus, there exists a minimal Dehn diagram D bordered by $wx^{-1}u^{-1}x$. It follows from the isoperimetry in G that $|D| \leqslant (|u| + |w| + 2|x|)/C$. As $|w| \leqslant |u|$ we have $|D| \leqslant |u|(4 + 2C'_c)/C$.

Do the same job with v and w, to get a diagram D' bordered by $v^{-1}y^{-1}wy$. Then paste these two diagrams along the w's, getting a diagram bordered by $v(xy)^{-1}u^{-1}(xy)$. Then transform this diagram into an annulus by gluing the two xy sides; this leads to a one-hole diagram bordered by u and v. The number of its faces is at most $(|u| + |v|)(4 + 2C'_c)/C$ and we conclude by setting $C' = C/(4 + 2C'_c)$ in case u and v are geodesic.



In case u is not geodesic, let u' be a geodesic word equal to u in G. We know there exists a Dehn diagram D_u bordered by uu'^{-1}, with $|D_u| \leqslant 2|u|/C$. Let D_v be a similar diagram for v. Let D be as above a one-hole minimal diagram bordered by u' and v', with $|D| \leqslant (|u| + |v|)/C'$ with C' as above. Then we can glue D_u and D_v to D along their common borders.



This leads to a diagram with at most $(|u| + |v|)/C' + 2(|u| + |v|)/C$ faces, and we conclude by re-setting C' to $1/(1/C' + 2/C)$. $\square$

B.3 Narrowness of diagrams

We now prove that diagrams (with or without holes) in a hyperbolic space are narrow (see section 1 for definitions).

Let $\alpha = 1/\log(1/(1-C'/\lambda))$ where C' is given by Proposition 50. (Recall λ is the maximal length of relators in the presentation of G.) Let $\lceil x \rceil$ denote the integer part of x plus one (such that $\lceil \log |D| \rceil = 1$ for $|D| = 1$).

Proposition 51 – *Let D be a minimal diagram with either 0 or 1 hole. Then D is $\lceil \alpha \log |D| \rceil$-narrow.*

Proof – Let D be a minimal Dehn diagram with 0 or 1 hole. Proposition 50 tells us that $C' |D| \leqslant |\partial D|$. Let n be the number of faces of D lying on the boundary. We have $|\partial D| \leqslant \lambda n$. Thus the proportion of faces of D lying on the boundary is at least C'/λ.

Let D' be the diagram D with the boundary faces removed. (In case D' is not connected, consider one of its connected component.) D' has at most one hole. D' is minimal as a subdiagram of a minimal diagram. Furthermore, we have $|D'| \leqslant |D| (1 - C'/\lambda)$. By the same argument, the proportion of boundary faces of D' is at least C'/λ, and after removing these faces we get a third diagram D'' with at most $|D| (1-C'/\lambda)^2$ faces. Repeating the argument yields the desired conclusion as D is exhausted after $\log |D| / \log(1/(1 - C'/\lambda))$ steps. $\square$

Proposition 52 – *Let D be a minimal n-hole diagram. Then D satisfies the isoperimetric inequality*

$$|\partial D| \geqslant C |D| - n\lambda (2 + 4\lceil \alpha \log |D| \rceil)$$

Proof –

Lemma – *Let D be a minimal n-hole Dehn diagram ($n \geqslant 1$). Either there exists a path in the 1-skeleton of D joining two holes, with length at most $\lambda(1 + 2\lceil \alpha \log |D| \rceil)$, or there exists a path in the 1-skeleton of D joining one hole with the exterior boundary, with length at most $\lambda(1/2 + \lceil \alpha \log |D| \rceil)$.*

Proof of the lemma – We work by induction on n. Set $e = \lceil \alpha \log |D| \rceil$.

Observe that a chain of N adjacent faces provides a path of length at most $N\lambda/2$ in the 1-skeleton between any two vertices of these faces.

For $n = 1$, the lemma is clear: by the last proposition, the diagram is e-narrow, thus the two components of the boundary are linked by a chain of at most $2e$ faces, providing a path of length at most λe.

Now suppose the lemma is true up to some $n \geqslant 1$, and let D be a $(n+1)$-hole Dehn diagram. For every hole i, let B_i be the set of faces of D lying at distance at most $2e + 1$ from the boundary of i.

Either, first, there are holes $i \neq j$ such that B_i and B_j have a common face or edge or vertex. This provides a chain of at most $4e + 2$ faces between the boundaries of holes i and j, thus a path of length at most $\lambda(2e + 1)$.

Or, second, the B_i's do not meet. Choose any hole i.

There can be holes in B_i, different from i, that can be filled in D. Define B'_i as B_i plus the interiors of these holes in D, such that all holes of B'_i are holes of D.

First, suppose that B'_i does not encircle any hole j of D other than i. As B_i is defined as the bowl of radius $2e+1$ around i in D, any face on the exterior boundary of B'_i is either a face at distance $2e+1$ from hole i, or a face on the boundary of D. But as B'_i is a one-hole Dehn diagram included in D, hence e-narrow by Proposition 51, not all faces of the exterior boundary of B'_i can be at distance $2e+1$ from i. That is, at least one face of the exterior boundary of B'_i is on the exterior boundary of D, hence a path of length at most $\lambda(e+1/2)$.

Second, imagine that B'_i encircles at least one hole $j \neq i$ of D. Consider the part D' of D comprised between B'_i and j, that is, the connected component of $D \setminus B'_i$ containing j. This is a diagram with at least one hole j (and maybe others), but as it does not contain i it has at most n holes. As D is minimal, D' is. By the induction hypothesis, either two holes in D' are at distance at most $\lambda(2e+1)$, in which case we are done, or one hole, say j, in D' is at distance at most $\lambda(e+1/2)$ of the exterior boundary of D'. But the exterior boundary of D' is part of the boundary of B'_i, any point of which is at distance at most $\lambda(e+1/2)$ of hole i. Thus i and j are linked by a path of length at most $\lambda(2e+1)$, which ends the proof of the lemma. □

Corollary of the lemma – *A minimal n-hole diagram can be made simply connected by cutting it along n curves of cumulated length at most $n\lambda(2\lceil\alpha\log|d|\rceil+1)$.*

The corollary of the lemma ends the proof of the proposition. □

Corollary 53 – *A minimal n-hole diagram D is $\lceil\alpha\log|D|\rceil + n(4\lceil\alpha\log|d|\rceil+2)$-narrow.*

Proof – Let D' be a simply connected Dehn diagram resulting from cutting D along curves of cumulated length at most $n\lambda(2\lceil\alpha\log|d|\rceil+1)$ (which run along at most $n(4\lceil\alpha\log|d|\rceil+2)$ faces as can immediately be seen on the proof above). Every face in the new diagram is at distance $\lceil\alpha\log|D|\rceil$ from the boundary of D' by Proposition 51. The boundary of D is a subset of the boundary of that of D', but by construction any face on the boundary of D' is at distance at most $n(4\lceil\alpha\log|d|\rceil+2)$ from the boundary of D. □

References

[A] G.N. Arzhantseva, *Generic properties of finitely presented groups and Howson's theorem*, Comm. Alg. **26**, No. 4 (1998), 3783–3792.

[AO] G.N. Arzhantseva, A.Yu. Ol'shanskii, *Generality of the class of groups in which subgroups with a lesser number of generators are free*, Mat. Zametki **59**, No. 4 (1996), 489–496; translation in Math. Notes **59**, No. 3-4 (1996) 350–355.

[BH] M.R. Bridson, A. Haefliger, *Metric Spaces of Non-Positive Curvature*, Grundlehren der mathematischen Wissenschaften **319**, Springer (1999).

[C] J.M. Cohen, *Cogrowth and Amenability of Discrete Groups*, J. Funct. Anal. **48** (1982), 301–309.

[Ch1] C. Champetier, *Propriétés statistiques des groupes de présentation finie*, J. Adv. Math. **116**, No. 2 (1995), 197–262.

[Ch2] C. Champetier, *Cocroissance des groupes à petite simplification*, Bull. London Math. Soc. **25**, No. 5 (1993), 438–444.

[Ch3] C. Champetier, *L'espace des groupes de type fini*, Topology **39**, No. 4 (2000), 657–680.

[GH] É. Ghys, P. de la Harpe, *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math. **83**, Birkhäuser (1990).

[Gri] R.I. Grigorchuk, *Symmetrical Random Walks on Discrete Groups*, in *Multicomponent Random Systems*, ed. R.L. Dobrushin, Ya.G. Sinai, Adv. Prob. Related Topics **6**, Dekker (1980), 285–325.

[Gro1] M. Gromov, *Hyperbolic Groups*, in *Essays in group theory*, ed. S.M. Gersten, Springer (1987), 75–265.

[Gro2] M. Gromov, *Asymptotic Invariants of Infinite Groups*, in *Geometric group theory*, ed. G. Niblo, M. Roller, Cambridge University Press, Cambridge (1993).

[Gro3] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*, Progress in Math. **152**, Birkhäuser (1999).

[Gro4] M. Gromov, *Random Walk in Random Groups*, Preprint IHÉS (2002).

[K1] H. Kesten, *Symmetric Random Walks on Groups*, Trans. Amer. Math. Soc. **92** (1959), 336–354.

[K2] H. Kesten, *Full Banach Mean Values on Countable Groups*, Math. Scand. **7** (1959), 146–156.

[LS] R.C. Lyndon, P.E. Schupp, *Combinatorial Group Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete **89**, Springer (1977).

[Ol] A.Yu. Ol'shanskii, *Almost Every Group is Hyperbolic*, Int. J. Algebra Comput. **2** (1992), No. 1, 1–17.

[Pap] P. Papasoglu, *An Algorithm Detecting Hyperbolicity*, in G. Baumslag (ed.) et al., *Geometric and Computational Perspectives on Infinite Groups*, DIMACS Ser. Discrete Math. Theor. Comput. Sci. **25** (1996), 193–200.

[S] A. Short et al., in *Group Theory from a Geometrical Viewpoint*, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991).

[SW] P. Scott, T. Wall, *Topological Methods in Group Theory*, in *Homological group theory*, ed. C.T.C. Wall, London Math. Soc. Lecture Notes Series **36** (1979), 137–203.

[W1] W. Woess, *Cogrowth of Groups and Simple Random Walks*, Arch. Math. (Basel) **41** (1983), 363–370.

[W2] W. Woess, *Random Walks on Infinite Graphs and Groups*, Cambridge Tracts in Mathematics **138**, Cambridge University Press (2000).

[Z] A. Żuk, *Property (T) and Kazhdan constants for discrete groups*, Preprint École normale supérieure de Lyon.

Contents

Introduction **1**

1 Definitions and notations **6**
1.1 Basics . . . 6
1.2 Growth, cogrowth, and gross cogrowth . . . 6
1.3 Diagrams . . . 9
1.4 Isoperimetry and narrowness . . . 10

2 The standard case: F_m **11**
2.1 Triviality for $d > 1/2$ . . . 12
2.2 Hyperbolicity for $d < 1/2$ . . . 12

3 Outline of the argument **18**
3.1 A basic picture . . . 19
3.2 Foretaste of the Axioms . . . 21

4 Axioms on random words implying hyperbolicity of a random quotient, and statement of the main theorem **21**
4.1 Asymptotic notations . . . 22
4.2 Some vocabulary . . . 22
4.3 The Axioms . . . 23
4.4 The Theorem . . . 25
4.5 On Axiom 4 . . . 25

5 Applications of the main theorem **27**
5.1 Satisfaction of the axioms . . . 28
5.1.1 The case of plain random words . . . 28
5.1.2 The case of random geodesic words . . . 30
5.1.3 The case of random reduced words . . . 34
5.2 Triviality of the quotient in large density . . . 34
5.2.1 The case of plain random words . . . 35
5.2.2 The case of random geodesic words . . . 36
5.2.3 The case of random reduced words . . . 36
5.3 Elimination of the virtual centre . . . 36
5.3.1 The case of plain or reduced random words . . . 37
5.3.2 The case of random geodesic words . . . 38

6 Proof of the main theorem **38**
6.1 On the lengths of the relators . . . 39
6.2 Combinatorics of Dehn diagrams of the quotient . . . 39
6.3 Coarsening of a Dehn diagram . . . 43
6.4 Graph associated to a decorated abstract Dehn diagram . . . 45
6.5 Elimination of doublets . . . 46
6.6 Pause . . . 50

6.7 Apparent length 51
6.8 The main argument 53
6.9 Non-elementarity of the quotient 60
6.9.1 Infiniteness 60
6.9.2 Non-quasi$\mathbb{Z}$ness 61

A Appendix: The Cartan-Hadamard-Gromov-Papasoglu theorem 64

B Appendix: Conjugacy and isoperimetry in hyperbolic groups 69
B.1 Conjugate words in G 69
B.2 One-hole diagrams 72
B.3 Narrowness of diagrams 73