École normale supérieure de Lyon
Unité de mathématiques pures et appliquées

Le Hasard et la Courbure

Yann Ollivier

Mémoire déposé en vue d'obtenir le diplôme
d'habilitation à diriger des recherches

présenté le 28 mai 2009 à l'École normale supérieure de Lyon

Jury :

M. Dominique Bakry
M. Persi Diaconis (rapporteur)
M. Étienne Ghys
M. Mikhail Gromov (rapporteur)
M. Michel Ledoux (rapporteur)
M. Robert J. McCann
M. Pierre Pansu

Le Hasard et la Courbure

Randomness and Curvature

Avant-propos

La démarche commune à la plupart des travaux présentés ici est l'étude géométrique d'objets quelconques, typiques, ou irréguliers. La typicité est obtenue par l'utilisation, à un endroit ou à un autre, d'ingrédients aléatoires. Cela peut signifier que l'objet a été obtenu par tirage aléatoire dans sa classe, par perturbation aléatoire autour d'un modèle mieux compris, ou bien qu'un objet est fixé mais suffisamment inconnu pour que le seul point d'attaque consiste à l'observer aléatoirement.

Un des objectifs de cette géométrie « synthétique » ou « robuste » est d'obtenir des arguments qui, lorsqu'ils s'appliquent à un espace, s'appliquent aussi bien à des espaces « proches ». Ces derniers peuvent être, par exemple, des espaces discrets comme un graphe, ou bien des variétés dont la géométrie a un grand nombre de fluctuations ou d'irrégularités à très petite échelle.

Le premier invariant géométrique que l'on rencontre en s'éloignant de l'espace euclidien est la notion de courbure, qui apparaîtra souvent dans ces pages. On peut sommairement diviser son influence en deux continents : celui de la courbure négative (ou majorée), et celui de la courbure positive (ou minorée). L'un comme l'autre seront aperçus ici, mais d'aucun des deux, je ne peux prétendre à une vue d'ensemble.

D'où « Le Hasard et la Courbure ».

Cette démarche sera appliquée à trois domaines bien différents. Dans le premier, il s'agit de *groupes aléatoires*, dont le comportement donne des indications sur ce que peut être un groupe quelconque, par opposition aux groupes classiques bien connus. Ici c'est la courbure négative qui domine : les groupes aléatoires sont *hyperboliques*, et beaucoup des propriétés qu'on leur connaît tournent autour de ce fait.

Dans le second, nous adopterons un point de vue géométrique sur les *chaînes de Markov*. On verra en particulier comment utiliser ces dernières pour définir une notion de *courbure (de Ricci)* sur des espaces métriques quelconques, qui permet d'étendre certaines propriétés classiques des variétés de courbure positive, comme la concentration de la mesure.

Enfin, la troisième partie, physique plus que mathématique, traite de *relativité générale* : l'équation d'Einstein lie la courbure au contenu en matière de l'espace-temps, et des fluctuations aléatoires à petite échelle, nulles en moyenne et non observées, peuvent avoir un effet non trivial sur la courbure à grande échelle de l'Univers. Cet effet physique de « matière apparente » est étudié dans différentes situations.

Comme le précisent les décrets, ce mémoire d'habilitation se compose d'un document de synthèse ainsi que d'un dossier de travaux scientifiques. Ce dernier reprend presque tous mes articles, y compris ceux de ma thèse (certains travaux ultérieurs en dépendant fortement), ainsi que le texte de mon petit livre sur les groupes aléatoires.

J'ai fait figurer, en tête du texte de synthèse, une table synoptique des principaux résultats, tentant de résumer chacun en une phrase. L'exercice n'est pas aussi réussi que je l'aurais souhaité, mais je l'espère utile.

Foreword

The common idea underlying the various works presented here is a geometric study of generic, typical, or irregular objects. Typicality is achieved through the use of random ingredients at one point or another. The object under scrutiny may have been picked at random in its class, or be a random perturbation of a smoother, more symmetric model, or just be a plain metric space with no particular features, for which random measurements provide the only reasonable approach.

One of the goals of this “coarse” or “robust” geometry is to develop geometric arguments that remain valid when considering objects that are “close” to a given one. The perturbed object might not be regular; typical examples include discrete spaces like graphs, or manifolds with many small fluctuations in their metric.

When departing from Euclidean space, the first geometric invariant encountered is curvature; this notion will pervade our work. Its influence can be broadly divided into two realms: that of positive curvature (or bounded below), and that of negative curvature (or bounded above), which entail very different behaviors. Glimpses of each will be given here, but I cannot pretend to a global view of either.

Hence “Randomness and Curvature”.

Three fairly different applications will be used to illustrate these principles. *Random groups* will come first. Their behavior hints at what a “generic” group looks like, as opposed to the more classical groups we all learn about. Random groups belong with negatively curved spaces: they are *hyperbolic*, and most of their known features arise from hyperbolicity.

Next, we will develop a geometric viewpoint on *Markov chains*, and see how random walk considerations lead to a notion of *(Ricci) curvature* on arbitrary metric spaces. Several classical properties of positively curved manifolds, such as concentration of measure, extend to this setting.

A bit of *general relativity* will come last; our treatment there will be physical rather than mathematical. The Einstein equation relates the matter content of spacetime to its curvature in a non-linear way, and small, unobserved fluctuations of matter may vanish on average, yet have a non-trivial effect on the large-scale curvature and dynamics of the Universe. This physical effect of an emerging “apparent matter” is investigated in a variety of situations.

As law and custom have it, this habilitation document consists of a survey together with a compilation of works. The latter comprises almost all my articles, including those from my PhD thesis (on which some later works heavily rely), as well as the text of my short book on random groups.

I decided to begin the survey with a synoptic table of results, each condensed to one sentence. This has proven harder to write than I expected; still, I hope it is not totally useless.

Curriculum vitæ sommaire

1978	Naissance.
1997–2001	Scolarité à l'École normale supérieure de Paris.
2001–2004	Thèse à l'Université Paris-Sud (Orsay), sous la direction de Misha Gromov et Pierre Pansu.
2004–. . .	Chargé de recherche à l'UMPA, École normale supérieure de Lyon.

Remerciements

Mes remerciements vont tout d'abord au CNRS, qui nous permet de travailler librement et dans de bonnes conditions. Je souhaite que cela soit le cas longtemps encore.

À l'UMPA, on a la chance de côtoyer chaque jour des personnalités d'une richesse exceptionnelle tant d'un point de vue humain que scientifique. Ajoutons que la communication en particulier entre différents sous-domaines des mathématiques, est inégalée ailleurs. Virtuellement tous les membres du laboratoire m'ont apporté à l'occasion leurs connaissances, leur compagnie ou les deux ; sans parler de l'aide de secrétaires remarquablement dévouées. Je veux remercier tous ces collègues, dont l'influence directe ou indirecte sur les travaux ici présentés est évidente.

Je tiens à saluer bien bas mes rapporteurs, Persi Diaconis, Misha Gromov (sans qui, tout bonnement, les objets et le point de vue mathématique utilisés ici n'existeraient pas) et Michel Ledoux, qui ont dû ingérer un lourd document en un temps assez court ; ainsi que les autres membres de mon jury, Dominique Bakry, Étienne Ghys, Robert McCann, Pierre Pansu, pour l'honneur qu'ils me font de s'intéresser à mon travail, parfois depuis fort longtemps. Car l'intérêt des collègues est une des gratifications essentielles de notre métier.

Enfin, merci à tous mes coauteurs, Fabrice, Dani, Pierre, Claire, Aldéric, Cédric, pour le plaisir de travailler ensemble ; que l'on me permette d'ajouter que certains sont bien plus que de simples collaborateurs.

Table des matières
Contents

Liste des travaux

reproduits dans ce mémoire

Table synoptique des résultats
Synoptic table of results

Théorie géométrique des groupes
Geometric group theory

Résultat 1.
Un quotient d'un groupe hyperbolique sans torsion par des éléments choisis au hasard dans une très grande boule, est encore un groupe hyperbolique sans torsion, non trivial en densité inférieure à 1/2.

Result 1.
A quotient of a torsion-free hyperbolic group by random elements chosen in a large enough ball, is still torsion-free hyperbolic, and non-trivial in density less than 1/2.

Sharp phase transition theorems for hyperbolicity of random groups, Theorem 3

Résultat 2.
Un quotient d'un groupe hyperbolique sans torsion par des mots aléatoires en les générateurs, est encore un groupe hyperbolique sans torsion, non trivial en densité inférieure à une certaine densité critique. La densité critique est donnée par l'exposant de retour en e de la marche aléatoire dans le groupe.

Result 2.
A quotient of a torsion-free hyperbolic group by randomly chosen words in the generators, is still torsion-free hyperbolic, and non-trivial up to some critical density. The critical density is given by the return exponent of the random walk in the group.

Sharp phase transition theorems for hyperbolicity of random groups, Theorem 4

Résultat 3.
Dans le théorème ci-dessus, l'exclusion de la torsion est nécessaire ; en présence de torsion, la densité critique peut être plus basse.

Result 3.
In the theorem above, torsion-freeness is necessary; in case of torsion, the critical density can be lower than expected.

Effondrement de quotients aléatoires de groupes hyperboliques avec torsion, Théorème 1

Résultat 4.
L'exposant de retour en e de la marche aléatoire dans un groupe hyperbolique sans torsion, n'est (presque) pas modifié par quotient aléatoire. En particulier, le trou spectral d'un groupe aléatoire est proche de celui d'un groupe libre.

Result 4.
The return exponent of the random walk in a torsion-free hyperbolic group, is (almost) invariant under a random quotient. In particular, the spectral gap in a random group is close to that in a free group.

Cogrowth and spectral gap of generic groups, Theorems 1 and 2

Résultat 5.
L'exposant de croissance d'un groupe hyperbolique sans torsion, n'est (presque) pas modifié par quotient aléatoire. En particulier, l'exposant de croissance d'un groupe aléatoire est très proche de celui d'un groupe libre.

Result 5.
The growth exponent of a torsion-free hyperbolic group, is (almost) invariant under a random quotient. In particular, the growth exponent of a random group is close to that of a free group.

Growth exponent of generic groups, Theorems 1 and 2

Résultat 6.
Dans un groupe hyperbolique, il est possible d'approximer l'exposant de croissance en se restreignant à une boule de rayon pas trop grand. En particulier, cet exposant de croissance est algorithmiquement calculable étant donné une présentation d'un groupe hyperbolique.

Result 6.
In a hyperbolic group, the growth exponent computed in a (not too) large enough ball is an explicit approximation of that of the whole group. In particular, this growth exponent is algorithmically computable given a presentation of a hyperbolic group.

Growth exponent of generic groups, Proposition 17 and Corollary 18

Résultat 7.
Dans un groupe hyperbolique, l'exposant de retour de la marche aléatoire peut être explicitement approché en se restreignant à des mots aléatoires assez longs. En particulier, cet exposant de retour est algorithmiquement calculable étant donné une présentation d'un groupe hyperbolique.

Result 7.
In a hyperbolic group, the return exponent of the random walk can be explicitly approximated using random words of a given, not too large length. In particular, this return exponent is algorithmically computable given a presentation of a hyperbolic group.

Cogrowth and spectral gap of generic groups, Proposition 8

Résultat 8.
Dans un groupe, il est bien connu qu'une inégalité isopérimétrique linéaire vraie sur des mots assez longs implique une inégalité isopérimétrique linéaire sur tous les mots (et donc l'hyperbolicité). Nous montrons que si l'on écrit l'inégalité isopérimétrique de la bonne manière, ceci peut se faire avec une perte arbitrairement faible dans la constante isopérimétrique.

Result 8.
Given a group presentation, it is known that a linear isoperimetric inequality valid for large enough words implies a linear isoperimetric inequality for all words (hence hyperbolicity). We show that if the isoperimetric inequality is written in a suitable way, this can be done with arbitrarily small loss in the isoperimetric constant.

Some small cancellation properties of random groups, Theorem 8

Résultat 9.
Un groupe aléatoire en densité $d < 1/2$ satisfait une inégalité isopérimétrique avec constante $1-2d$. En particulier on a la δ-hyperbolicité avec $\delta \leqslant 4\ell/(1-2d)$.

Result 9.
A random group at density $d < 1/2$ satisfies an isoperimetric inequality with constant $1-2d$. In particular, the δ-hyperbolicity constant satisfies $\delta \leqslant 4\ell/(1-2d)$.

Some small cancellation properties of random groups, Theorem 2 and Corollary 3

Résultat 10.
Dans un groupe aléatoire en densité $d < 1/5$, l'algorithme de Dehn pour le problème du mot est valable ; il ne l'est pas si $d > 1/5$.

Result 10.
In a random group at density $d < 1/5$, the Dehn algorithm for the word problem terminates, whereas it does not if $d > 1/5$.

Some small cancellation properties of random groups, Theorem 6

Résultat 11.
Un groupe aléatoire en densité $d < 1/5$ ne satisfait pas la propriété T de Kazhdan.

Result 11.
Random groups at density $d < 1/5$ do not have Kazhdan's property T.

Cubulating groups at density $1/6$, Corollary 51

Résultat 12.
Un groupe aléatoire en densité $d < 1/6$ agit librement et cocompactement sur un complexe cubique CAT(0) et possède la propriété de Haagerup.

Result 12.
Random groups at density $d < 1/6$ act freely and cocompactly on a CAT(0) cube complex and have the Haagerup property.

Cubulating groups at density 1/6, Theorem 62 and Corollary 56

Résultat 13.
La technique de « présentations graphiques » de Gromov permet de construire des groupes ayant la propriété (T) et qui sont, au choix, non hopfiens, non co-hopfiens, ou dont le groupe d'automorphismes extérieurs est infini.

Result 13.
Gromov's "graphical presentation" tool allows to construct groups with property (T) which are, respectively, not Hopfian, not co-Hopfian, or with infinite outer automorphism group.

Kazhdan groups with infinite outer automorphism group, Theorems 1.2, 1.3 and 1.4

Chaînes de Markov, concentration, courbure de Ricci
Markov chains, concentration, Ricci curvature

Résultat 14.
On peut définir une courbure de Ricci sur les espaces métriques mesurés, même discrets ; cette courbure est positive si les petites boules sont plus proches (en distance de transport) que leurs centres. Cette notion redonne la courbure de Ricci sur les variétés riemanniennes, est compatible avec celle de Bakry–Émery, et attribue une courbure positive à des espaces tels que le cube discret.

Result 14.
One can define a notion of Ricci curvature for metric measure spaces, including discrete spaces; this curvature is positive when small balls are closer (in transportation distance) than their centers are. This notion gives back the usual Ricci curvature of a Riemannian manifold and is consistent with Bakry–Émery theory; such spaces as the discrete cube are positively curved.

Ricci curvature of Markov chains on metric spaces, Definition 3 and Examples 7, 8 and 11

Résultat 15.
Comme dans le cas riemannien, la stricte positivité de la courbure de Ricci discrète permet :
– de contrôler le diamètre (th. de Bonnet–Myers) ;
– de contrôler le trou spectral du laplacien (th. de Lichnerowicz) ;
– d'obtenir la concentration de la mesure (th. de Lévy–Gromov) ;
– d'obtenir la contraction de gradient par le flot de la chaleur et une inégalité de Sobolev logarithmique (th. de Bakry–Émery).

Result 15.
As in the Riemannian case, positive Ricci curvature in the sense above implies:
– a diameter control (Bonnet–Myers thm.);
– a spectral gap estimate (Lichnerowicz thm.);
– concentration of measure (Lévy–Gromov thm.);
– gradient contraction by the heat equation, and a logarithmic Sobolev inequality (Bakry–Émery thms.).

Ricci curvature of Markov chains on metric spaces, Propositions 23 and 30 and Theorems 33, 44 and 45

Résultat 16.
L'utilisation de fonctions de Green discrètes permet de définir quantitativement une notion de voisinage dans un graphe ou une chaîne de Markov. Appliquée au graphe des liens internes de Wikipédia, cette stratégie permet de renvoyer automatiquement une liste d'articles de Wikipédia « sur le même sujet » qu'un article donné.

Result 16.
Discrete Green functions allow to define a quantitative notion of neighborhood in a graph or a Markov chain. When tested on the graph of internal links of Wikipedia, this provides a fully automated way of listing articles "related to" a given Wikipedia article.

Finding related pages using Green measures: An illustration with Wikipedia

Résultat 17.
Dans une population finie se reproduisant par reproduction sexuée avec choix aléatoire du partenaire dans toute la population, le brassage des gènes est exponentiellement rapide, mais l'équilibre atteint diffère du modèle d'une population infinie par un terme décroissant comme l'inverse de la taille de la population.

Result 17.
In a finite population evolving through sexual reproduction with random choice of the mate in the population, genes get mixed exponentially fast, but the final equilibrium differs from the ideal model of an infinite population by an amount inversely proportional to population size.

Rate of convergence of crossover operators, Theorems 4 and 8

Physique statistique relativiste
General relativistic statistical physics

Résultat 18.
Des ondes gravitationnelles actuellement indétectables pourraient avoir un effet très important sur la dynamique globale de l'univers. Par contre, les effets globaux de fluctuations de la densité de matière seraient plus faibles.

Result 18.
Currently undetectable gravitational waves could have considerable effect on the dynamics of the Universe. In contrast, fluctuations of the density of matter seem to have a much smaller effect.

Large-scale non-linear effects of fluctuations in relativistic gravitation
& Multiscale cosmological dynamics

Résultat 19.
Des imprécisions dans l'observation d'un trou noir donnent à croire qu'il est entouré de matière d'énergie négative.

Result 19.
Imprecision when observing a black hole yields the impression that it is surrounded with "apparent matter" of negative energy.

Observing a Schwarzschild black hole with finite precision

Présentation

Comme signalé dans l'avant-propos, beaucoup des travaux de ce mémoire utilisent la notion de courbure, soit directement, soit pour s'en inspirer. Il est donc utile de passer un peu de temps à examiner cette notion.

Notre objectif n'est pas de reproduire un cours de géométrie riemannienne, cours auquel nous référons le lecteur pour le détail des définitions. Nous nous attarderons néanmoins sur l'intuition que l'on peut avoir des différents objets en jeu, intuition souvent passée sous silence. On pourra consulter par exemple [Car92] pour une première approche, ou [Ber03] pour un impressionnant survol des différents aspects de la géométrie riemannienne.

1 Les courbures en géométrie riemannienne

Variétés riemanniennes. Rappelons que l'archétype d'une variété riemannienne est une surface plongée dans l'espace euclidien. Plus généralement, toute variété (lisse) peut être vue comme un ensemble $X \subset \mathbb{R}^p$ tel qu'en tout point, il existe un sous-espace affine de dimension N dans $\mathbb{R}^p$ qui coïncide avec X au premier ordre. Ce sous-espace est appelé *espace tangent* T_xX au point $x \in X$ considéré, et N est la *dimension* de X.

Remarquons que si $c(t)$ est une courbe lisse dans X, alors la dérivée $\mathrm{d}c(t)/\mathrm{d}t$ est un vecteur tangent à M en $c(t)$.

Une variété est *riemannienne* si elle est en outre équipée d'une *métrique riemannienne*, c'est-à-dire, en chaque point x, d'une forme quadratique définie positive sur l'espace tangent T_xX. Cela peut être par exemple, si $X \subset \mathbb{R}^p$, la restriction à T_xX d'une structure euclidienne sur $\mathbb{R}^p$.

Une telle forme quadratique permet d'attribuer une norme à tout vecteur tangent, et, par intégration, de définir la longueur d'une courbe dans X. La distance (dans X) entre deux points de X est alors définie comme l'infimum des longueurs des courbes les joignant, ce qui fait de X un espace métrique.

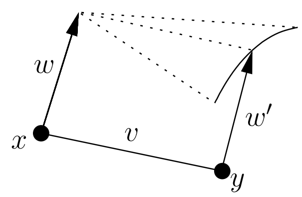
On supposera toujours que X est connexe ainsi que complet pour cette métrique.

Une *géodésique* est une courbe γ dans X telle que, étant donné deux points de γ assez proches l'un de l'autre, γ réalise la distance entre ces deux points. Localement, de telles courbes existent toujours. De plus, étant donné un point $x \in X$ et un vecteur tangent $v \in T_xM$, il existe exactement une géodésique issue de x ayant v comme vitesse initiale, qu'on appellera géodésique issue de v.

On appellera *extrémité* de v, et on notera $\exp_x v$, le point obtenu en suivant cette géodésique pendant un temps unité.

Transport parallèle. Supposons que l'on ait deux points x et y très proches dans une variété riemannienne. A-t-on un moyen de comparer un vecteur tangent en x à un vecteur tangent en y, qui vivent *a priori* dans deux espaces différents ? C'est ce qu'autorise la notion de *transport parallèle*.

Puisque x et y sont très proches, on peut supposer que y est l'extrémité d'un vecteur tangent v en x. Maintenant, supposons qu'on ait un autre vecteur tangent w en x, très petit lui aussi, et supposons, pour simplifier, que w est orthogonal à v. Il existe alors un vecteur tangent particulier w' en y : celui dont l'extrémité est la plus proche de celle de w, parmi tous les vecteurs tangents en y orthogonaux à v (si on supprime cette condition d'orthogonalité, bien sûr, on peut trouver un vecteur tangent en y dont l'extrémité est exactement la même que celle de w, mais ce vecteur « revient vers x »). C'est le meilleur candidat à être le vecteur tangent en y qui soit « le même » que w.



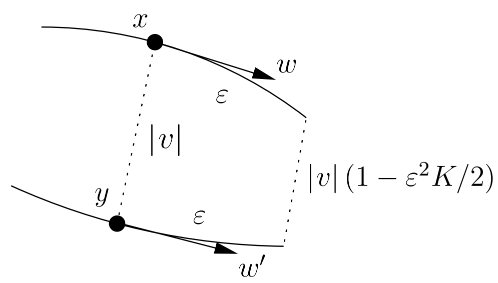
On appelle ce vecteur w' le *transporté parallèle* de w le long de v (plus exactement, le transport parallèle est la linéarisation de cette opération, c'est-à-dire qu'on suppose w très petit et qu'on étend ensuite par linéarité). On peut lever la condition d'orthogonalité à v en décidant que par définition, le transporté parallèle de v le long de lui-même est le vecteur tangent en y à la géodésique issue de v.

Plus généralement, on peut définir le transport parallèle de w le long d'une courbe issue de x en décomposant la courbe en intervalles très petits et en faisant des transports parallèles successifs le long de ces intervalles.

Courbure sectionnelle et courbure de Ricci. Passons maintenant à la courbure, en commençant par la première de ses variantes, la courbure sectionnelle.

Reprenons notre point x, notre vecteur tangent v d'extrémité y, notre vecteur tangent w en x et son transporté parallèle w'. Si la variété est l'espace euclidien, les extrémités x' et y' de w et w' forment, avec x et y, un rectangle. Mais, dans une variété riemannienne quelconque, ce n'est plus le cas.

En effet, les deux géodésiques issues de w et w' vont, en présence de courbure, avoir tendance à se rapprocher ou à s'éloigner. Ainsi sur la sphère (courbure positive), deux méridiens issus de points très proches de l'équateur se rencontrent aux pôles. Comme w et w' sont parallèles, cet effet est du second ordre en la distance parcourue.



Considérons ainsi la distance entre les points situés à distance ε de x ou y sur les géodésiques issues respectivement de w et w'. Cette distance serait égale à $|v|$ dans le cas euclidien, et on utilise la différence pour mesurer une courbure.

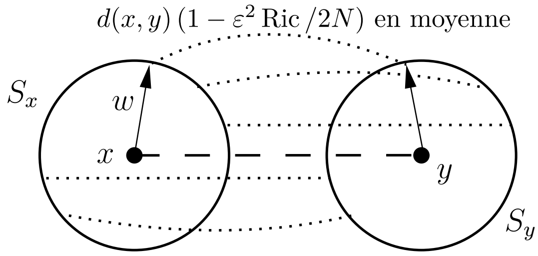
Définition 1 (Courbure sectionnelle).
Soit (X,d) une variété riemannienne. Soient v et w des vecteurs tangents unitaires en un point $x \in X$. Soient $\varepsilon, \delta > 0$. Soit y l'extrémité de δv et soit w' le transporté parallèle de w de x vers y. Alors

$$d(\exp_x \varepsilon w, \exp_y \varepsilon w') = \delta \left(1 - \frac{\varepsilon^2}{2} K(v,w) + O(\varepsilon^3 + \varepsilon^2 \delta)\right)$$

lorsque $(\varepsilon, \delta) \to 0$. La quantité $K(v,w)$ ainsi définie est appelée courbure sectionnelle *dans les directions (v,w).*

La *courbure de Ricci*, elle, ne dépend que d'un seul vecteur tangent v et est obtenue en moyennant sur toutes les directions w.

Définition 2 (Courbure de Ricci).
Soit x un point d'une variété riemannienne de dimension N. Soit v un vecteur tangent en x. On appelle courbure de Ricci le long de v, *la quantité* $\mathrm{Ric}(v)$ *égale à N fois la moyenne de $K(v,w)$ où la moyenne est prise sur w parcourant la sphère unité de l'espace tangent en x.*



Le facteur N provient de la définition traditionnelle de la courbure de Ricci comme une trace, qui fournit une somme sur une base plutôt qu'une moyenne sur la sphère. De plus, il se trouve que $\mathrm{Ric}(v)$ est une forme quadratique en v, que nous noterons donc plutôt $\mathrm{Ric}(v,v)$.

On peut reformuler cette définition de la manière suivante.

Corollaire 3.
Soit v un vecteur tangent unitaire en un point x d'une variété riemannienne. Soient $\varepsilon, \delta > 0$ et soit y l'extrémité de δv.

Soit S_x l'ensemble des extrémités des vecteurs tangents en x de norme ε, et, de même, S_y l'ensemble des extrémités de la sphère de rayon ε dans l'espace tangent en y. Alors, si l'on envoie S_x sur S_y par transport parallèle, en moyenne les points sont déplacés d'une distance

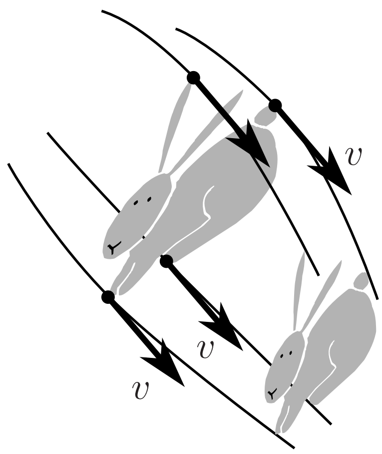
$$\delta\left(1 - \frac{\varepsilon^2}{2N}\operatorname{Ric}(v,v) + O(\varepsilon^3 + \varepsilon^2\delta)\right)$$

lorsque $(\varepsilon, \delta) \to 0$.

Si les sphères sont remplacées par des boules, le facteur $\frac{\varepsilon^2}{2N}$ devient $\frac{\varepsilon^2}{2(N+2)}$.

En particulier, la courbure de Ricci est positive si « les boules sont plus proches que leurs centres ». On verra que cette propriété s'adapte très bien à des espaces beaucoup plus généraux.

Mentionnons au passage une autre manière, plus dynamique, de visualiser la courbure de Ricci. Soit à nouveau un vecteur tangent unitaire v en un point x d'une variété riemannienne X. Soit C un petit voisinage de x, que l'on peut choisir d'une forme quelconque. Pour tout point z de C, considérons la géodésique z_t issue de z et ayant v pour vitesse initiale (où v a été préalablement amené en z par transport parallèle). On a vu qu'en moyenne, ces géodésiques se rapprochent ou s'éloignent selon le signe de la courbure. Faisons maintenant « glisser » l'ensemble C le long de ces géodésiques ; plus précisément, soit C_t l'ensemble $\{z_t,\ z \in C\}$.



Alors on a

$$\operatorname{vol} C_t = \operatorname{vol} C\,(1 - \tfrac{t^2}{2}\operatorname{Ric}(v,v))$$

à des termes d'ordre supérieur près en t et en la taille de C. (Notons que la dérivée de $\operatorname{vol} C_t$ en $t = 0$ est nulle car les géodésiques considérées ont des vitesses initiales parallèles.) Autrement dit, la courbure de Ricci contrôle la contraction des volumes par le flot géodésique.

Les signes de la courbure. Souvent en géométrie riemannienne, lorsque l'on veut travailler en courbure négative on doit supposer que toutes les courbures sectionnelles $K(v, w)$ sont négatives, et pour travailler en courbure positive on doit supposer que la courbure de Ricci $\mathrm{Ric}(v, v)$ est positive pour tout v (ce qui est plus faible que de supposer la positivité de tous les $K(v, w)$).

Nous n'avons certainement pas l'ambition de donner une vue d'ensemble des applications de la courbure en géométrie riemannienne, aussi donnerons-nous un exemple très simple de chaque cas. En courbure de Ricci positive on a le théorème de Bonnet–Myers, qui affirme que si une variété est plus courbée que la sphère, alors son diamètre est inférieur.

Théorème 4 (Bonnet–Myers).
Soit X une variété riemannienne de dimension N. Soit $\inf \mathrm{Ric}(X)$ l'infimum de tous les $\mathrm{Ric}(v, v)$ pour v vecteur tangent unitaire.

Soit S^N la sphère unité de dimension N dans $\mathbb{R}^{N+1}$. Si $\inf \mathrm{Ric}(X) \geqslant \inf \mathrm{Ric}(S^N)$, alors $\operatorname{diam} X \leqslant \operatorname{diam} S^N$.

Beaucoup de théorèmes en courbure de Ricci positive prennent la forme d'un théorème de comparaison avec la sphère. Nous reviendrons plus loin, en particulier, sur le phénomène de concentration de la mesure.

En courbure négative, mentionnons le théorème de Cartan–Hadamard, qui affirme que toute l'information topologique est contenue dans le groupe fondamental de la variété. Ceci est complètement faux en courbure positive, puisque par exemple une sphère est simplement connexe.

Théorème 5 (Cartan–Hadamard).
Soit X une variété riemannienne de dimension N telle que $K(v, w) \leqslant 0$ pour tous vecteurs tangents v, w. Alors le revêtement universel de X est homéomorphe à $\mathbb{R}^N$.

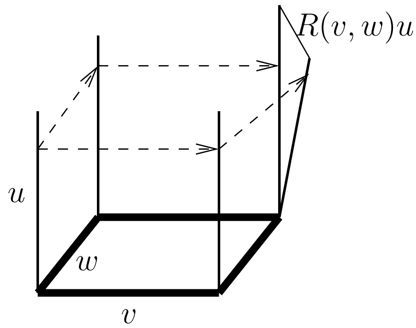
Une autre propriété du plan hyperbolique et plus généralement des variétés (simplement connexes) de courbure sectionnelle strictement négative, est que la surface d'une très grande boule est comparable à son volume (alors que par exemple, dans l'espace euclidien, la surface d'une très grande boule croît comme la puissance $2/3$ de son volume). De telles *inégalités isopérimétriques linéaires* seront très importantes dans notre étude des groupes aléatoires.

Le tenseur de courbure de Riemann. Continuons dans notre parcours des notions de la géométrie riemannienne. Le *tenseur de courbure de Riemann* prend trois vecteurs tangents u, v, w en un point x, et renvoie un nouveau vecteur tangent noté $R(v, w)u$. La notation se justifie par le fait que, à v et w fixés, l'application $u \mapsto R(v, w)u$ est une application linéaire de l'espace tangent en x dans lui-même. Avant de définir le tenseur de Riemann, discutons plus avant la notion de transport parallèle.

Lorsqu'on a deux vecteurs tangents v, w en un point x, on peut transporter v le long de w, ou transporter w le long de v, et les extrémités des vecteurs obtenus sont a priori différentes. En fait, lorsque v et w sont petits, ces deux opérations coïncident au premier ordre où il n'est pas trivial qu'elles coïncident (précisément, à l'ordre $|v|\,|w|$ quand v, w sont petits). On appelle cette propriété l'*absence de torsion*, elle affirme qu'à cet ordre « les parallélogrammes se referment ».

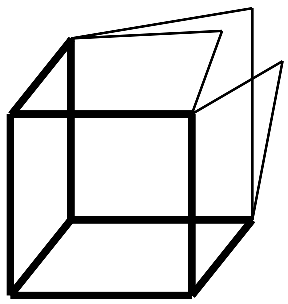
Nous avons implicitement utilisé cette propriété plus haut, puisque la définition de la courbure faisait intervenir une quantité d'ordre supérieur, précisément $|v|\,|w|^2$.

Essayons maintenant avec trois vecteurs u, v, w. Gardons le même parallélogramme construit sur v et w, et essayons de faire le transport parallèle de u le long du chemin vw, et le long du chemin wv. Cette fois-ci, les deux résultats diffèrent, et leur différence définit la *courbure de Riemann* $R(v, w)u$, qui est à nouveau un vecteur tangent.



Ceci prouve qu'un cube ne se referme pas, contrairement à un parallélogramme.

Étudions ce problème de plus près. On a vu qu'atteindre le sommet opposé du cube produisait des résultats différents selon qu'on suivait le chemin vwu ou wvu, la différence étant précisément $R(v, w)u$. Il y a au total six chemins possibles. Néanmoins, leurs extrémités coïncident deux à deux : en effet, grâce à l'absence de torsion on a bien sûr $vw = wv$ mais aussi $uvw = uwv$ qui exprime juste l'absence de torsion en l'extrémité de u. Autrement dit, « les faces latérales du cube se referment ». Sur les six chemins possibles, on n'obtient donc que trois points d'arrivée.[1]



[1] Avec Bruno Sévennec, nous avons pu transposer cet argument en une démonstration rigoureuse très courte, en interprétant le petit cube dans un quotient bien choisi d'un espace de jets.

Ces trois points forment un petit triangle. On a vu que l'un des côtés de ce triangle était $R(v,w)u$, et par symétrie les autres sont $R(w,u)v$ et $R(u,v)w$. La relation de Chasles dans ce triangle s'écrit donc

$$R(u,v)w + R(v,w)u + R(w,u)v = 0$$

qui est la *première identité de Bianchi*, découverte par Ricci.

Volume riemannien. Pour clore ce tour informel des bases de la géométrie riemannienne, mentionnons que la donnée d'une métrique riemannienne permet de mesurer non seulement les distances mais aussi les volumes. En effet, on peut décréter que le volume d'un petit parallélépipède orthonormé est 1.

Ceci définit une forme volume sur une variété riemannienne, que l'on note souvent $\sqrt{\det g}$ si g est la métrique. La raison de cette notation est la suivante. En un point donné, la métrique g est une forme bilinéaire sur l'espace tangent E, c'est-à-dire une application de E vers E^*. Passons aux puissances extérieures. Le déterminant $\det g = \bigwedge^N g$ est une application de $\bigwedge^N E$ dans $\bigwedge^N E^* \simeq \left(\bigwedge^N E\right)^*$. Autrement dit, $\det g$ est une forme bilinéaire sur $\bigwedge^N E$, qui est en outre définie positive si g l'est. Sa racine $\sqrt{\det g}$ est donc une norme sur $\bigwedge^N E$, c'est-à-dire exactement une forme volume.

Contrairement à ce qu'on lit souvent, la notation $\sqrt{\det g}$ n'a ainsi rien d'abusif...

Applications. Passons maintenant en revue les différents domaines d'application intervenant dans ce mémoire. D'abord, les groupes aléatoires, dont la propriété centrale est la δ-hyperbolicité (« courbure négative discrète ») et ses différentes conséquences géométriques. En ce sens, au moins dans le monde des groupes, la courbure négative est « générique ». Ensuite, nous passerons à la notion de courbure de Ricci discrète, qui permet de généraliser un certain nombre de propriétés des variétés à courbure de Ricci positive, en particulier la concentration de la mesure. Enfin, nous décrirons des travaux de physique : en relativité générale, la courbure « à grande échelle » d'une variété irrégulière diffère de la moyenne de la courbure, et ceci créée un effet physique de « matière apparente » que nous décrirons dans certaines situations, et qui peut évoquer le problème de la matière noire.

2 Courbures discrètes I : courbure sectionnelle négative

2.1 La δ-hyperbolicité

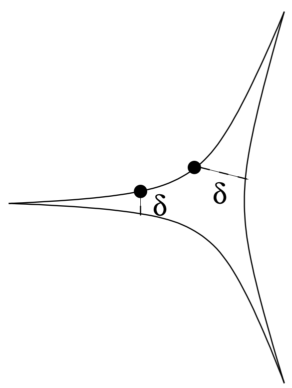
La δ-hyperbolicité, notion remontant au moins à Rips mais considérablement développée par Gromov [Gro87], est une des principales propriétés utilisées pour généraliser la courbure sectionnelle négative, et s'est révélée particulièrement fructueuse pour l'étude de certains groupes discrets, exemples sur lesquels nous nous concentrerons. Les principaux résultats sur ces *groupes hyperboliques* sont déjà présents dans [Gro87], tandis que le livre [GhH90] y donne une introduction. Le texte très complet [BH99] traite des différents aspects des notions de courbure négative dans les espaces métriques, incluant le cas des groupes.

Soit X un espace métrique. Un *segment géodésique* dans X est un plongement isométrique d'un intervalle réel $[0;\ell]$ dans X. On supposera ici que l'espace X est tel que, pour toute paire de points x, y dans X avec $d(x,y) = \ell$, il existe un segment géodésique de longueur ℓ joignant x et y (pas forcément unique). C'est le cas, par exemple, d'une variété riemannienne, ou bien d'un graphe si l'on considère que chaque arête est un segment d'une longueur donnée.

Un *triangle* dans X est la donnée de trois points de X, ainsi que de trois segments géodésiques les reliant deux à deux qu'on appellera *côtés* du triangle.

Définition 6 (Triangles δ-fins).
Soit δ un nombre positif. On dit qu'un triangle est δ-fin *si, pour tout point sur un côté du triangle, ce point est à distance au plus δ de l'un des deux autres côtés.*

Intuitivement, cela signifie que le triangle est très aplati, et que l'espace laissé au milieu est de largeur environ δ.



Définition 7 (δ-hyperbolicité).
Soit δ un nombre positif. On dit que l'espace métrique X est δ-hyperbolique si tout triangle de X est δ-fin.

La définition est motivée par le fait que le plan hyperbolique standard est δ-hyperbolique. Il en est de même de toute variété riemannienne simplement connexe à courbure sectionnelle majorée par $-K$, $K > 0$.

Un des intérêts de la notion de δ-hyperbolicité est sa robustesse : intuitivement, elle n'est pas affectée par des modifications de l'espace à des échelles très petites devant δ, et peut donc se concevoir comme une courbure sectionnelle négative « à grande échelle ». En particulier, un espace proche d'un espace hyperbolique, en un sens que l'on peut préciser (quasi-isométrie), est encore hyperbolique.

Étudions maintenant une classe d'espaces qui ne sont pas des variétés mais pour lesquels la notion de δ-hyperbolicité est pertinente : les groupes discrets.

2.2 Quelques notions de géométrie des groupes

Groupes hyperboliques. Considérons un groupe G engendré par un nombre fini d'éléments $a_1, \ldots, a_m$ et leurs inverses. On peut voir G comme un graphe dont les sommets sont tous les éléments de G, avec une arête entre x et y si $x = ya_i^{\pm 1}$ pour un certain a_i. Ce graphe est appelé le *graphe de Cayley* de G, et dépend bien sûr de la famille génératrice choisie. La connexité de ce graphe traduit le fait que les a_i engendrent G.

On peut faire du graphe de Cayley un espace métrique en décidant que chaque arête est de longueur 1 et que la distance entre deux sommets est la longueur du plus court chemin les joignant.

Définition 8 (Groupes hyperboliques).
Un groupe muni d'une famille génératrice $(a_1, \ldots, a_m)$ *est* hyperbolique *si son graphe de Cayley est δ-hyperbolique pour un certain $\delta \geqslant 0$.*

C'est une propriété non triviale que l'hyperbolicité ne dépend pas de la famille génératrice utilisée pour définir le graphe de Cayley. Une manière de le voir consiste à démontrer l'équivalence (non triviale) entre cette définition et l'inégalité isopérimétrique linéaire décrite ci-dessous, qui elle est facilement invariante par changement de famille génératrice.

Présentations de groupes. D'un point de vue combinatoire, le plus simple des groupes engendrés par m éléments est le *groupe libre* à m générateurs, noté F_m. Ce groupe consiste en l'ensemble des mots sur l'alphabet à $2m$ lettres $a_1, \ldots, a_m, a_1^{-1}, \ldots, a_m^{-1}$ avec la condition que ces mots sont *réduits* i.e. ne contiennent pas une lettre suivie immédiatement de son inverse. La multiplication de ce groupe est la concaténation des mots avec éventuelle simplification des couples $a_i^{\pm 1} a_i^{\mp 1}$ qui pourraient apparaître à la jonction ; l'élément neutre est le mot vide.

En fait tout groupe G engendré par m éléments $a_1, \ldots, a_m$ et leurs inverses peut être vu comme un quotient du groupe libre F_m : en effet étant donné un mot dans ce groupe libre, on peut former le produit des éléments de G correspondants et on obtient un élément de G ; cette application est surjective puisque les $a_i^{\pm 1}$ engendrent le groupe.

Cela correspond au fait qu'il y a toujours plus de « règles de calcul » dans G que dans le groupe libre. Ici par *règle de calcul* on entend une égalité $w_1 = w_2$ où w_1 et w_2 sont deux mots en les $a_i^{\pm 1}$. (Par exemple la règle de calcul $a_1 a_2 = a_2 a_1$ est satisfaite lorsque a_1 et a_2 commutent, et si ces deux éléments engendrent le groupe, alors tout le groupe est commutatif.)

Il existe alors un moyen simple de produire un groupe satisfaisant une règle de calcul $w_1 = w_2$ donnée. Il suffit de quotienter le groupe libre F_m par le plus petit sous-groupe normal contenant l'élément $w_1 w_2^{-1}$. Dans le groupe G ainsi obtenu, noté $G = \langle\, a_1, \ldots, a_m \mid w_1 = w_2 \,\rangle$, le mot $w_1 w_2^{-1}$ représente l'identité, et donc les mots w_1 et w_2 représentent le même élément.

Plus généralement, donnons-nous un ensemble (fini ou infini) R de mots du groupe libre F_m. Le groupe noté $\langle\, a_1, \ldots, a_m \mid \{r = e\}_{r \in R} \,\rangle$, ou plus simplement $\langle\, a_1, \ldots, a_m \mid R \,\rangle$, est le groupe $G = F_m / \langle R \rangle$ où $\langle R \rangle$ est le plus petit sous-groupe normal de F_m contenant R. Les éléments de R seront appelés *relateurs*. Tout groupe de type fini admet une telle présentation.

Inégalité isopérimétrique linéaire et diagrammes de van Kampen. Lorsqu'un groupe G est présenté par $G = \langle\, a_1, \ldots, a_m \mid R \,\rangle$, tout mot w en les générateurs représentant l'élément neutre de $G = F_m / \langle R \rangle$ est par construction un élément de $\langle R \rangle$. En conséquence il s'écrit comme un produit de conjugués d'éléments de R ou de leurs inverses :

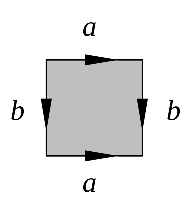
$$w = e \text{ dans } G \Leftrightarrow w = \prod u_i r_i^{\pm 1} u_i^{-1},\ r_i \in R$$

l'égalité ayant lieu en tant que mots dans le groupe libre.

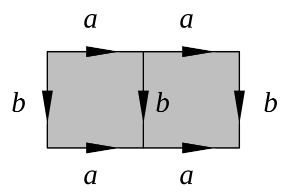
Une question naturelle qui se pose alors est : en présence d'un mot dont on sait qu'il représente l'élément neutre, combien de relateurs r_i comporte, au minimum, une telle décomposition ? Un des premiers résultats de la théorie des groupes hyperboliques est le suivant (on renvoie à [BH99] pour une preuve).

Proposition 9.
Un groupe $G = \langle\, a_1, \ldots, a_m \mid R \,\rangle$ est hyperbolique si et seulement s'il existe une constante C telle que pour tout mot w de longueur L représentant l'élément neutre de G, w peut s'écrire comme un produit d'au plus $C.L$ conjugués de relateurs.

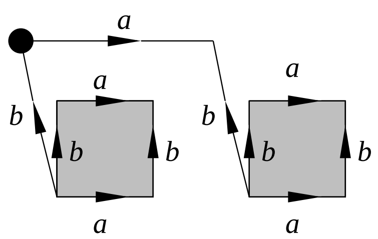
Il existe une interprétation géométrique de ces produits de conjugués de relateurs. Étant donné une présentation de groupe $G = \langle\, a_1, \ldots, a_m \mid R \,\rangle$, pour chaque relateur $r \in R$ (supposé réduit), de longueur L_r, on définit un *relateur géométrique* comme un disque bordé par L_r arêtes, chaque arête portant un générateur a_i comme suit : la k-ième arête porte le générateur correspondant à la k-ième lettre de r (on met une orientation inverse sur l'arête si la k-ième lettre de r est a_i^{-1}). Voici par exemple le relateur géométrique associé au relateur $aba^{-1}b^{-1}$.



On peut former des « puzzles » avec ces relateurs géométriques, où l'on s'autorise à recoller deux relateurs géométriques le long d'arêtes identiques (on peut aussi utiliser les relateurs inverses). Cela définit un *diagramme de van Kampen*. Van Kampen a prouvé qu'un mot (réduit) représente l'élément neutre dans G si et seulement si ce mot peut être lu sur le bord d'un diagramme de van Kampen. Voici par exemple une preuve que si a et b commutent, alors a^2 et b commutent.



Les diagrammes de van Kampen sont liés aux produits de conjugués de relateurs de la manière suivante : choisir un point-base dans le diagramme, suivre un chemin jusqu'à un premier relateur, faire le tour du relateur, revenir au point-base, suivre un chemin jusqu'à un deuxième relateur, en faire le tour, revenir au point-base, etc. On décrit alors un mot de la forme $\prod u_i r_i^{\pm 1} u_i^{-1}$, les u_i correspondant aux trajets entre le point-base et les relateurs. Par exemple, en refermant le diagramme ci-dessous on retrouve celui donné ci-dessus.



Dans les groupes hyperboliques, on sait donc qu'étant donné un mot w représentant l'élément neutre, on peut trouver un diagramme de van Kampen ayant ce mot comme bord, et dont le nombre de faces croît au plus linéairement en la taille de w. Intuitivement, un tel diagramme a donc une aire proportionnelle à son périmètre, comme un disque dans le plan hyperbolique, d'où l'appellation *inégalité isopérimétrique linéaire*.

L'inégalité isopérimétrique linéaire et les diagrammes de van Kampen seront nos principaux outils dans l'étude des groupes aléatoires.

2.3 Le monde des groupes aléatoires

L'une des raisons avancées par Gromov [Gro87] pour s'intéresser aux groupes hyperboliques est que « la plupart » des groupes le sont. La formalisation de cette affirmation est l'objet de l'étude des groupes aléatoires. Mon texte de survol *A January 2005 invitation to random groups*, reproduit dans ce mémoire, contient une introduction au sujet, aussi cette présentation sera-t-elle brève. On pourra aussi consulter [Ghy03].

Puisqu'on sait que tout groupe admet une présentation $G = \langle\, (a_i)_{i\in I} \mid R\,\rangle$, on peut choisir un groupe au hasard en tirant au hasard une telle présentation. On peut alors se demander si telle ou telle propriété du groupe est très probable ou non.

La manière dont on tire la présentation au hasard constitue un *modèle* de groupe aléatoire. La plupart des modèles qui ont été étudiés jusqu'ici utilisent un nombre fini, fixé, de générateurs $a_1, \ldots, a_m$ et spécifient ensuite comment tirer un ensemble aléatoire R de mots en ces générateurs. Un modèle particulièrement intéressant est le *modèle à densité* [Gro93], dans lequel un paramètre $d \in [0;1]$ permet de contrôler la quantité de relateurs que l'on place dans R, ce qui permet d'étudier précisément l'influence du nombre de relateurs sur les caractéristiques du groupe. Plus précisément, le modèle à densité dépend de deux paramètres, la densité $d \in [0;1]$ ainsi qu'un entier ℓ qui contrôle la longueur des relateurs. Ce dernier paramètre ℓ est supposé grand, ce qui permet d'obtenir des propriétés dont la probabilité tend vers 1 quand $\ell \to \infty$.

Notons qu'on peut toujours supposer qu'une présentation de groupe ne contient que des relateurs réduits, puisque toute paire $a_i^{\pm 1} a_i^{\mp 1}$ apparaissant dans un relateur peut être supprimée sans changer le groupe. Maintenant, le nombre N_ℓ de mots réduits d'une certaine longueur ℓ sur l'alphabet $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$ est $N_\ell = (2m)(2m-1)^{\ell-1}$. Le modèle à densité consiste à prendre un nombre de relateurs égal à une certaine puissance $d \in [0;1]$ de ce nombre total de relateurs possibles.

Modèle à densité.
Choisir un nombre d entre 0 *et* 1*. Se donner une longueur de mots ℓ très grande. Tirer un ensemble de relateurs R en tirant $(N_\ell)^d$ fois de suite (indépendamment, avec ou sans remise) un mot réduit au hasard uniformément parmi les N_ℓ mots réduits de longueur ℓ possibles. Un* groupe aléatoire à densité d *est le groupe $G = \langle\, a_1, \ldots, a_m \mid R\,\rangle = F_m/\langle R\rangle$ ainsi obtenu.*

L'intérêt de cette manière de fixer la taille de R est démontré par le théorème suivant, dû à Gromov [Gro93].

Théorème 10 (Transition de phase pour les groupes aléatoires).
Soit G un groupe aléatoire à densité d. Si $d < 1/2$, la probabilité que G soit infini et hyperbolique tend vers 1 *lorsque $\ell \to \infty$. Si $d > 1/2$, le groupe G est soit $\{e\}$ soit $\mathbb{Z}/2\mathbb{Z}$, avec probabilité tendant vers* 1 *quand $\ell \to \infty$.*

On trouvera une démonstration de ce résultat à la fin de *A January 2005 Invitation to Random Groups*, mais esquissons tout de même une preuve de la partie $d > 1/2$ du théorème. L'idée, qui fait bien ressortir le pourquoi de la densité, repose sur le *principe des tiroirs probabiliste* : si l'on place au hasard beaucoup plus que $\sqrt{N}$ objets dans N tiroirs, alors très probablement deux objets sont placés le même tiroir. Par exemple, dans une classe de $20 = \left\lceil \sqrt{365} \right\rceil$ élèves, il y a déjà plus de 40% de chances que deux d'entre eux aient la même date de naissance.

En densité $d > 1/2$, on a donc très probablement tiré deux fois le même mot dans l'ensemble de relateurs R. Cela n'est pas très intéressant mais, a fortiori, on a aussi très probablement tiré dans R deux mots qui ne diffèrent que par la première lettre. Supposons par exemple que $r_1 = a_1 w$ et $r_2 = a_2 w$ ont été placés dans R, où w est un mot de longueur $\ell - 1$. Par définition du groupe aléatoire G, r_1 et r_2 représentent tous les deux l'élément neutre de G. Mais cela implique immédiatement que $a_1 = a_2$ dans G. En fait, quand $\ell \to \infty$, très probablement chaque générateur a_i devient égal à tous les autres ainsi qu'à leurs inverses, et le groupe G ne peut alors être que $\{e\}$ ou $\mathbb{Z}/2\mathbb{Z}$ (ce dernier cas correspondant à ℓ pair).

En densité plus petite que $1/2$, l'argument consiste à construire des diagrammes de van Kampen et à démontrer qu'ils satisfont une inégalité isopérimétrique linéaire. L'idée est que, lorsqu'on fabrique un diagramme de van Kampen en recollant des relateurs aléatoires, chaque recollement « coûte » un facteur $1/(2m-1)$ en probabilité. Ceci permet de borner la quantité de recollements possibles lorsqu'on fabrique les diagrammes de van Kampen, et de prouver qu'une certaine longueur reste forcément sur le bord du diagramme, qui satisfait donc une inégalité isopérimétrique.

Intéressons-nous à des généralisations du Théorème 10. Ce dernier affirme qu'un groupe aléatoire, autrement dit un quotient aléatoire d'un groupe libre, est hyperbolique. On peut se demander si un quotient aléatoire d'un groupe hyperbolique reste hyperbolique, autrement dit, si l'hyperbolicité est stable en plus d'être générique.

Les deux résultats ci-dessous constituent les théorèmes principaux du long article *Sharp phase transition theorems for hyperbolicity of random groups*. On rappelle qu'un groupe est *sans torsion* s'il n'existe pas d'élément x (à part e) et d'entier $n > 1$ avec $x^n = e$. Un groupe hyperbolique sans torsion est en outre dit *non élémentaire* s'il n'est ni $\{e\}$ ni $\mathbb{Z}$.

Théorème 11.
Soit G_0 un groupe hyperbolique sans torsion et non élémentaire. Fixons une famille génératrice finie de G_0, et soit B_ℓ l'ensemble des éléments de G_0 de taille au plus ℓ par rapport à cette famille génératrice.

Soit $0 \leqslant d \leqslant 1$. Soit $R \subset G_0$ un ensemble obtenu en tirant au hasard $(\#B_\ell)^d$ fois de suite un élément de B_ℓ (uniformément, avec ou sans remise). Soit $G = G_0/\langle R \rangle$ le quotient aléatoire obtenu.

- *Si $d < 1/2$, la probabilité que G soit hyperbolique non élémentaire tend vers 1 quand $\ell \to \infty$.*
- *Si $d > 1/2$, alors $G = \{e\}$ avec probabilité tendant vers 1 quand $\ell \to \infty$.*

Autrement dit, dans un groupe hyperbolique on peut « tuer » beaucoup d'éléments choisis au hasard. Bien sûr, quand G_0 est un groupe libre on retrouve le théorème précédent (au remplacement près, indolore, de la boule par la sphère).

Étant donné une famille génératrice dans un groupe hyperbolique, il peut être plus commode de tirer des mots aléatoires en les générateurs plutôt qu'un élément dans la boule B_ℓ, la mesure uniforme sur B_ℓ étant plus difficile à simuler. C'est l'objet du théorème suivant.

Théorème 12.
Soit G_0 un groupe hyperbolique sans torsion et non élémentaire, engendré par des éléments $a_1, \ldots, a_m$.

Soit $0 \leqslant d \leqslant 1$. Soit W_ℓ l'ensemble des $(2m)^\ell$ mots de longueur ℓ sur l'alphabet $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$. Soit R un ensemble obtenu en tirant $(2m)^{d\ell}$ fois de suite un élément de W_ℓ (uniformément, avec ou sans remise). Soit $G = G_0/\langle R\rangle$ le quotient aléatoire obtenu.

Alors il existe un $d_{G_0} \in]0;1[$ tel que

- *Si $d < d_{G_0}$, la probabilité que G soit hyperbolique non élémentaire tend vers 1 quand $\ell \to \infty$.*
- *Si $d > d_{G_0}$, alors $G = \{e\}$ ou $\mathbb{Z}/2\mathbb{Z}$ avec probabilité tendant vers 1 quand $\ell \to \infty$.*

De plus, d_{G_0} peut être explicitement décrit comme l'exposant de retour en 0 de la marche aléatoire dans G_0 :

$$d_{G_0} = -\lim_{\substack{t\to\infty \\ t \text{ pair}}} \frac{1}{t} \log_{2m} \Pr(w_t =_{G_0} e)$$

où w_t est un mot aléatoire de longueur t en les $a_i^{\pm 1}$.

Décrivons très rapidement les autres propriétés des groupes aléatoires que j'ai pu obtenir. Comme la densité critique dépend de l'exposant de retour en 0 de la marche aléatoire dans le groupe de départ, il est naturel de se demander comment cette quantité est affectée par le fait de prendre un quotient aléatoire. J'ai montré dans *Cogrowth and spectral gap of generic groups* qu'en fait cette quantité n'est presque pas modifiée. Dans le cas du groupe libre, on peut interpréter ce résultat comme suit, en voyant les relations d'un groupe comme des « ponts » reliant l'élément neutre à certains éléments du groupe libre : si l'on fait une marche aléatoire dans un arbre, mais en ajoutant au hasard un grand nombre de ponts de longueur nulle entre l'origine et des sommets lointains choisis au hasard (ceci de manière covariante), soit tous les points sont reliés à l'origine par une succession de ponts (ceci correspond à une densité supérieure à $1/2$), soit la marche aléatoire ne voit essentiellement aucune différence.

J'ai montré dans *Growth exponent of generic groups* que l'exposant de croissance (taux de croissance exponentiel asymptotique du nombre de points dans une grande boule) d'un groupe aléatoire est en fait très proche de celui d'un groupe libre. Ceci répondait en partie à une question de Grigorchuk et de la Harpe [GrH97]. Un résultat intermédiaire notable est que l'exposant de croissance d'un groupe hyperbolique (non aléatoire) est algorithmiquement calculable.

Dans la note *Collapsing of random quotients of hyperbolic groups with torsion* j'ai montré que la présence d'éléments de torsion peut modifier la valeur de la densité critique des quotients aléatoires d'un groupe donné.

Je me suis aussi intéressé aux propriétés combinatoires des groupes aléatoires. Dans l'article *Some small cancellation properties of random groups* je montre, entre autres, qu'en densité inférieure à $1/5$ est satisfaite une propriété classique (l'algorithme de Dehn) utilisée pour résoudre le problème du mot. Un des outils est une version très fine du théorème de Cartan–Hadamard–Gromov pour les groupes hyperboliques, qui affirme qu'on peut vérifier l'inégalité isopérimétrique linéaire dans un groupe en connaissant seulement une partie finie du groupe ; j'ai raffiné ce résultat pour rendre la perte dans les constantes arbitrairement faible (au lieu d'un facteur 10^{10}), ce qui était indispensable pour mes applications.

Dans *On a small cancellation theorem of Gromov*, je donne une preuve détaillée d'une affirmation de Gromov, la « petite simplification à graphe » [Gro03], qui permet de produire des groupes dont le graphe de Cayley contient (presque injectivement) un graphe fini prescrit.

J'ai aussi entamé en 2004 une collaboration avec D. Wise (McGill University, Toronto) qui a produit deux articles. Nous avons développé de nouvelles techniques de construction de groupes ayant la propriété (T), qui sont relativement souples et permettent d'imposer d'autres conditions à volonté ; cela a permis de construire dans l'article *Kazhdan groups with infinite outer automorphism group* de nouveaux groupes répondant à des questions anciennes. Nous avons aussi étudié, dans le preprint *Cubulating groups at density* 1/6, certaines propriétés géométriques des groupes aléatoires, et en particulier nous avons montré un théorème de géométrisation pour les groupes aléatoires en densité $< 1/6$: ces groupes agissent essentiellement sur des complexes cubiques à courbure négative, et possèdent la propriété de Haagerup. Ces méthodes permettent aussi d'exclure la propriété (T) en densité $< 1/5$ (elle est connue en densité $> 1/3$).

Le texte *A January 2005 Invitation to Random Groups* tente de faire le point sur l'ensemble des propriétés connues des groupes aléatoires et donne un certain nombre de questions ouvertes sur le sujet.

3 Courbures discrètes II : courbure de Ricci positive

3.1 Courbure de Ricci discrète

Du point de vue de la physique statistique, il n'y a pas grande différence entre un système de n particules dont l'énergie totale est exactement E, et un système de n particules indépendantes ayant chacune une énergie moyenne E/n, du moins lorsque n est grand. Pour des particules libres où E est la somme des carrés des vitesses, dans le premier cas l'espace de configuration est une sphère dans l'espace des vitesses, tandis que dans le second cas, c'est tout l'espace des vitesses, mais muni d'une mesure gaussienne (la distribution de Maxwell–Boltzmann $\exp(-E/kT)$ associée à cette énergie E). En particulier, si on tire beaucoup de points au hasard suivant cette gaussienne, ils vont dessiner un ensemble proche d'une sphère. Un troisième modèle, discret celui-là, consisterait à donner à chaque particule une énergie 0 ou $2E/n$ en tirant à pile ou face, auquel cas l'espace de configuration serait le cube $\{0, 2E/n\}^n$.

On peut donc se demander si ces espaces ont quelque chose de commun du point de vue géométrique. Par exemple, une sphère étant l'archétype d'un espace de courbure positive, on peut se demander si l'espace euclidien, muni d'une mesure gaussienne, peut être qualifié d'espace à courbure positive.

Une théorie allant dans ce sens a été développée avec succès par Bakry et Émery [BE84, BE85]. Elle permet d'attribuer une courbure de Ricci à une variété munie d'une mesure (plus exactement à un espace muni d'un *processus de diffusion*, dont la mesure considérée est la distribution invariante), de sorte que l'espace gaussien acquière une courbure de Ricci positive. Néanmoins, l'utilisation des diffusions rend très délicate son application à des espaces discrets, comme notre troisième exemple, et, en ce sens, elle n'est pas robuste par passage à un espace « proche ». Les extensions de la théorie de Bakry–Émery proposées depuis (par exemple celle développée indépendamment dans [Stu06, LV, Oht07], voir aussi [RS05, OV00]) souffrent du même problème.

Il apparaît ainsi nécessaire de développer une notion robuste de courbure de Ricci « à une certaine échelle ». Cette notion devrait être compatible avec celle de Bakry–Émery, facile à appliquer sur des espaces discrets, permettre de généraliser des théorèmes connus en courbure de Ricci positive, et capturer de manière robuste les propriétés communes que la physique statistique observe pour la sphère ou l'espace gaussien. Ce dernier point nous amènera à parler de *concentration de la mesure*, application sur laquelle nous insisterons.

La notion proposée transpose directement la définition 2 ou plutôt son corollaire 3. Il s'agit de comparer la distance entre des petites boules à la distance entre leurs centres. La courbure de Ricci sera positive si *les boules sont plus proches que leurs centres*. Par analogie avec la définition 2 qui fait intervenir une mesure (on prend une moyenne sur une sphère tangente), il sera commode de voir une boule comme une mesure de masse 1 autour d'un point. Pour définir une distance entre boules, on peut alors simplement utiliser la *distance de transport* (ou de Wasserstein, ou Monge–Kantorovich–Rubinstein) entre mesures, notion bien connue [Vil03] définie comme suit.

Définition 13 (Distance de transport).
Soient μ_1, μ_2 *deux mesures de masse unité dans un espace métrique* (X, d)*. Un* plan de transfert *de* μ_1 *vers* μ_2 *(aussi appelé* couplage*) est une mesure* ξ *sur* $X \times X$ *telle que* $\int_y \mathrm{d}\xi(x,y) = \mathrm{d}\mu_1(x)$ *et* $\int_x \mathrm{d}\xi(x,y) = \mathrm{d}\mu_2(y)$*. (Ainsi* $\mathrm{d}\xi(x,y)$ *représente la quantité de masse déplacée de* x *en* y*.)*

La distance de transport L^1 *entre* μ_1 *et* μ_2*, notée* W_1*, est la meilleure distance moyenne réalisable :*

$$W_1(\mu_1, \mu_2) := \inf_{\xi \in \Pi(\mu_1, \mu_2)} \iint d(x,y)\,\mathrm{d}\xi(x,y)$$

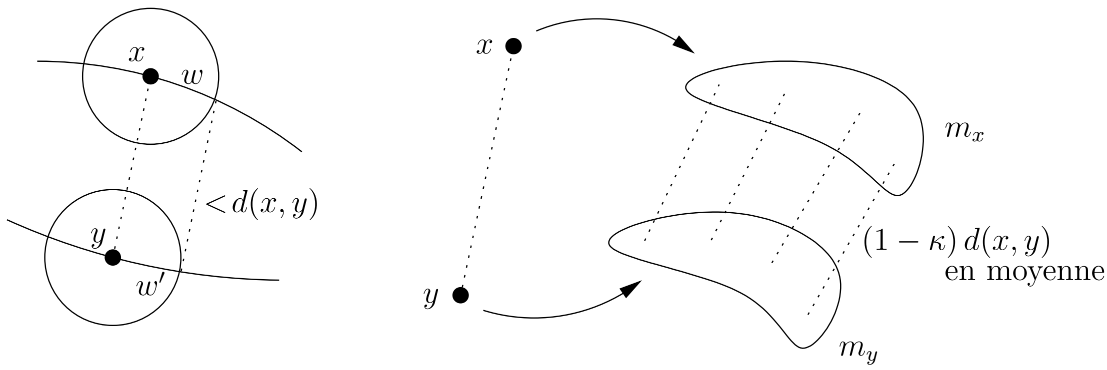
où $\Pi(\mu_1, \mu_2)$ *est l'ensemble des plans de transfert de* μ_1 *vers* μ_2*.*

En général W_1 peut être infini et est donc une semi-distance. L'inégalité triangulaire utilise le dit *lemme de recollement* pour les couplages, ce qui techniquement nécessite d'imposer que l'espace métrique X soit polonais (séparable, complet). W_1 est une véritable distance si on se restreint à l'ensemble des mesures μ ayant un *premier moment fini*, c'est-à-dire telles que $\int d(o,x)\,\mathrm{d}\mu < \infty$ pour une certaine origine $o \in X$ que l'on peut choisir arbitrairement.

On va maintenant utiliser la distance de transport entre des petites boules pour définir une courbure de Ricci. La notion pertinente de « petite boule » dépend de la situation. Par exemple, dans un graphe il est naturel de prendre des boules de rayon 1, tandis que sur une variété on prendra des boules arbitrairement petites. Cela permet de définir une notion de courbure de Ricci « à une certaine échelle » selon la taille des boules utilisées. Nous supposerons ainsi que, pour chaque point dans un espace X, on a choisi une mesure m_x sur X, de masse 1, qui jouera le rôle d'une petite boule autour de x. (La définition 15 permet de choisir un tel système de boules de manière souvent pertinente.)

Définition 14 (Courbure de Ricci discrète).
Soit (X, d) *un espace métrique. On suppose que pour chaque* $x \in X$*, est donnée une mesure de probabilité* m_x *sur* X*. Soient* x *et* y *deux points distincts de* X*. La* courbure de Ricci discrète *le long de* xy *est la quantité* $\kappa(x,y)$ *définie par la relation*

$$W_1(m_x, m_y) = (1 - \kappa(x,y))\, d(x,y)$$



Dans une variété riemannienne, la notion de courbure de Ricci était définie le long d'un vecteur tangent. Ici dans un espace métrique, le mieux que l'on puisse faire pour remplacer un vecteur tangent est d'utiliser une paire de points. On verra plus loin (proposition 19) qu'il suffit de calculer $\kappa(x,y)$ pour des paires de points suffisamment proches.

Les hypothèses techniques nécessaires au bon fonctionnement de cette définition sont les suivantes : (X,d) doit être un espace polonais, et chaque m_x doit avoir un premier moment fini (voir ci-dessus).

Remarquons que la donnée des $(m_x)_{x\in X}$ définit exactement le noyau de transition d'une chaîne de Markov. Notre définition peut être considérée comme une version métrique des coefficients ergodiques habituels (définis en utilisant la distance de variation totale des mesures). En fait, Dobrushin [Dob70] utilisait une notion très similaire pour étudier des systèmes de spins, et notre définition peut être considérée comme la convergence de travaux de Dobrushin et ses successeurs [Dob70, DS85, Dob96, BD97] sur les chaînes de Markov d'une part, et du courant plus géométrique ou analytique initié par Bakry et Émery.

Pour des travaux utilisant des idées proches, voir par exemple [RS05, Jou07, Oli, DGW04].

Exemples. Passons en revue des exemples d'application de cette définition, en commençant bien sûr par les variétés riemanniennes. Pour cela, il faut dans chaque cas choisir la famille de mesures $(m_x)_{x\in X}$ de manière appropriée. La manière la plus simple est la suivante.

Définition 15 (Marche aléatoire de pas ε).
Soit (X,d,μ) un espace métrique mesuré ; on suppose que les boules dans X sont de mesure finie et que $\operatorname{Supp}\mu = X$. *Fixons $\varepsilon > 0$. La* marche aléatoire de pas ε *sur X consiste, étant donné un point de départ x, à faire un saut aléatoire dans la boule de rayon ε autour de x avec probabilité proportionnelle à μ. Autrement dit, on pose* $m_x = \mu_{|B(x,\varepsilon)}/\mu(B(x,\varepsilon))$.

En considérant la marche aléatoire de pas ε très petit dans une variété riemannienne, on retrouve, à normalisation près, la courbure de Ricci ordinaire. La proposition suivante est une variante du corollaire 3 dont notre définition était inspirée.

Exemple 16 (Variétés riemanniennes).
Soit (X,d) une variété riemannienne lisse. Considérons la marche aléatoire de pas ε, pour ε assez petit. Soient $x,y\in X$ deux points assez proches. et soit v le vecteur tangent unitaire en x pointant vers y. Alors

$$\kappa(x,y)=\frac{\varepsilon^2\,\mathrm{Ric}(v,v)}{2(N+2)}+O(\varepsilon^3+\varepsilon^2 d(x,y))$$

La normalisation en ε^2 traduit le fait que la différence entre une variété et l'espace euclidien est du second ordre.

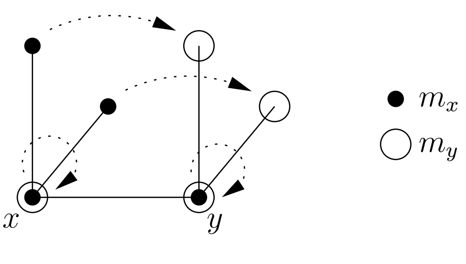
Voici un exemple qui n'est pas une variété, mais qui est proche d'un espace euclidien et dont on attend donc que la courbure soit nulle.

Exemple 17 ($\mathbb{Z}^N$ et $\mathbb{R}^N$).
Soit m la marche aléatoire de pas 1 sur la grille $\mathbb{Z}^N$ munie de sa métrique de graphe. Alors pour tous points $x,y\in\mathbb{Z}^N$, la courbure de Ricci discrète le long de xy est nulle.

Cette exemple se généralise à toute métrique et à toute marche aléatoire sur $\mathbb{Z}^N$ ou $\mathbb{R}^N$ qui soient invariantes par translation. Par exemple, le réseau triangulaire standard dans le plan est de courbure de Ricci discrète nulle.

L'exemple discret le plus intéressant est sans doute le cube. Comme nous l'avons déjà mentionné, la géométrie du cube en tant qu'espace métrique mesuré est assez semblable à celle de la sphère. Nous invitons le lecteur à refaire en détail l'exemple suivant, qui montre comment notre définition peut être calculée en pratique sur un espace discret.

Exemple 18 (Cube discret).
Soit le cube discret $\{0,1\}^N$, muni de sa métrique L^1 et de la mesure de probabilité uniforme. Considérons la marche aléatoire de pas 1. Alors la courbure de Ricci discrète d'une paire de points voisins x,y est $\kappa(x,y)=\frac{2}{N+1}$.



Le lecteur aura noté que l'estimation de $\kappa(x,y)$ dans cet exemple ne concerne que les paires de points voisins, et que de même, pour les variétés riemanniennes, nous avions supposé $d(x,y)$ très petit. En géométrie riemannienne, la courbure est une quantité locale, dont le contrôle permet ensuite d'obtenir des informations globales. La proposition suivante montre que, dans un espace *géodésique*, il en est de même de la courbure de Ricci discrète.

Proposition 19 (Espaces géodésiques).
Soit (X,d) un espace métrique α-géodésique, c'est-à-dire que pour tout couple de points $(x,y)\in X\times X$, il existe un entier n et une suite de points $x_0=x, x_1,\ldots,x_n=y$ avec $d(x,y)=\sum d(x_i,x_{i+1})$ et $d(x_i,x_{i+1})\leqslant\alpha$.

Alors, l'inégalité $\kappa(x,y)\geqslant\kappa$ pour tout couple de points (x,y) avec $d(x,y)\leqslant\alpha$, implique la même inégalité pour tout couple de points $(x,y)\in X\times X$.

Par exemple, un graphe est 1-géodésique et une variété riemannienne est α-géodésique pour tout α ; dans les deux cas, cela résulte de la construction même de la distance. Cette proposition est très simple à démontrer mais extrêmement utile dans les applications.

L'exemple suivant relie notre définition de la courbure de Ricci discrète à la notion de δ-hyperbolicité mentionnée dans la partie précédente. Bien que la courbure de Ricci négative ne soit pas très utile en pratique, il est agréable que les définitions soient compatibles.

Exemple 20 (Groupes hyperboliques).
Soit X le graphe de Cayley d'un groupe hyperbolique non élémentaire par rapport à une certaine famille génératrice. Soit k un entier assez grand et considérons la marche aléatoire sur X dont un pas consiste à faire k pas de la marche aléatoire simple par rapport à cette famille génératrice. Soient $x,y\in X$. Alors $\kappa(x,y)=-\frac{2k}{d(x,y)}(1-o(1))$ lorsque k et $d(x,y)$ tendent vers l'infini.

Remarquons que $-2k/d(x,y)$ est la plus petite valeur possible de κ pour une marche aléatoire dont les pas sont de taille au plus k.

Un point du cahier des charges pour la courbure de Ricci discrète était d'être compatible avec la théorie de Bakry–Émery. Rappelons que l'exemple le plus simple de cette dernière est le processus d'Ornstein–Uhlenbeck sur $\mathbb{R}$ ou $\mathbb{R}^N$, qui est le processus le plus naturel dont la gaussienne soit la mesure invariante. C'est un brownien modifié par une force de rappel linéaire vers l'origine, i.e. la solution de l'équation différentielle stochastique $\mathrm{d}X_t=\sqrt{2}\,\mathrm{d}B_t-X_t\,\mathrm{d}t$ associée à l'opérateur de diffusion $Lf=\Delta f-x\cdot\nabla f$. Plus généralement :

Exemple 21 (Courbure de Ricci d'après Bakry et Émery).
Soit X une variété riemannienne de dimension N et soit F un champ de vecteurs tangents sur X. Considérons l'opérateur différentiel

$$L:=\Delta+F\cdot\nabla$$

naturellement associé à l'équation différentielle stochastique

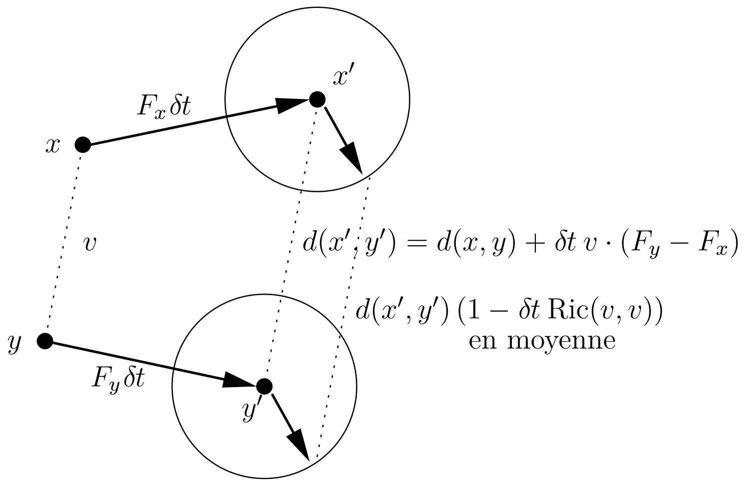
$$\mathrm{d}X_t=\sqrt{2}\,\mathrm{d}B_t+F\,\mathrm{d}t$$

où B_t est le mouvement brownien standard sur la variété riemannienne X. La courbure de Ricci de cet opérateur au sens de Bakry–Émery, appliquée à un vecteur v, est $\mathrm{Ric}(v,v)-v\cdot\nabla_v F$.

Considérons le schéma d'Euler suivant pour l'approximation au temps δt de ce processus. Partant d'un point x, on définit la mesure m_x en suivant le flot du champ F pendant un temps δt, puis en sautant aléatoirement dans une boule de rayon $\sqrt{2(N+2)\delta t}$ autour du point obtenu.

Soient $x, y \in X$ avec $d(x,y)$ assez petit, et soit v le vecteur tangent unitaire en x pointant vers y. Alors

$$\kappa(x,y) = \delta t \left(\mathrm{Ric}(v,v) - v \cdot \nabla_v F + O(d(x,y)) + O(\sqrt{\delta t}) \right)$$



Expliquons les normalisations. Sauter dans une boule de rayon ε engendre une variance $\varepsilon^2 \frac{1}{N+2}$ dans une direction donnée, tandis que le brownien standard a par définition une variance dt par unité de temps dans une direction donnée. En conséquence, la discrétisation correcte au temps δt exige de sauter dans une boule de rayon $\varepsilon = \sqrt{2(N+2)\delta t}$. Par ailleurs, le générateur infinitésimal du mouvement brownien est $\frac{1}{2}\Delta$ (laplacien des probabilistes) et non Δ, ce qui explique le facteur $\sqrt{2}$ supplémentaire.

L'origine du terme $-v \cdot \nabla_v F$ pour la courbure de Ricci au sens de Bakry–Émery apparaît clairement avec notre définition : cette quantité exprime la variation de la distance de deux points proches sous le flot de F. (En particulier, la partie antisymétrique de ∇F engendre une isométrie infinitésimale.)

Autres exemples. La courbure de Ricci discrète est facilement calculable et positive dans de nombreux autres exemples, comme les distributions binomiales ou multinomiales ou encore la mesure de Poisson (dans une limite convenable).

Mentionnons en particulier l'exemple du modèle d'Ising : la courbure de Ricci discrète est positive si et seulement si le *critère de Dobrushin* classique est satisfait, et les théorèmes généraux sur la courbure de Ricci discrète (concentration, inégalité de Sobolev logarithmique), appliqués à ce cas particulier, semblent être comparables à ceux de la littérature sur le sujet. L'article [Dob70] où ce critère est introduit est justement celui qui a renouvelé l'intérêt des mathématiciens pour les distances de transport et leur a donné le nom de *distances de Vasershtein*, qu'elles ont conservé à la transcription près.

3.2 Concentration de la mesure et courbure de Ricci

Quelques espaces concentrés. Passons maintenant aux théorèmes que l'on peut démontrer en utilisant la courbure de Ricci discrète. Plutôt que de les énoncer tous, nous nous attarderons sur l'un d'entre eux, un théorème de *concentration de la mesure*. Sur ce sujet, on pourra consulter l'introduction [Sch01], les ouvrages de référence [Led01, Mas07], ou encore le chapitre $3\frac{1}{2}$ de [Gro99] pour un point de vue très géométrique.

Revenons aux trois espaces mentionnés dans l'introduction à propos de N particules en physique statistique : la sphère, l'espace gaussien, et le cube discret.

Commençons par ce dernier. Soit $X = \{P, F\}^N$ l'espace des résultats de N tirages à pile ou face. Considérons la fonction $f : X \to \mathbb{R}$ égale à la proportion de « pile ». Un des résultats les plus fondamentaux des probabilités est que, pour N grand, la fonction f est presque toujours proche de $1/2$; en outre, les écarts à $1/2$ sont d'ordre $1/\sqrt{N}$ et à peu près gaussiens. On peut formaliser cette dernière affirmation en évaluant la mesure des points où l'écart est plus grand que t :

$$\mu\left(\{x \in X,\ \left|f(x) - \tfrac{1}{2}\right| \geqslant t\}\right) \leqslant 2 \exp -\frac{t^2}{2D^2}$$

où μ est la mesure de probabilité uniforme sur X et où $D = 1/\sqrt{N}$ représente l'écart-type.

En fait, et c'est l'intérêt des résultats de concentration, cette propriété est loin de se limiter à une seule fonction f. Plus exactement, le même résultat est vrai pour toute fonction f telle que, si l'on change un « pile » en « face » ou inversement, la valeur de f n'est modifiée que d'au plus $1/N$.

Reformulons cette dernière propriété en munissant notre espace X de la métrique suivante : la distance entre deux points $x = (x_1, \ldots, x_N)$ et $x' = (x'_1, \ldots, x'_N)$ de X est $1/N$ fois le nombre de différences entre x et x'. Ainsi, une fonction $f : X \to \mathbb{R}$ est 1-lipschitzienne si et seulement si elle varie d'au plus $1/N$ quand on change un « pile » en « face » ou inversement. Nous renvoyons à [Led01] pour une démonstration du résultat suivant, désormais folklorique.

Théorème 22.
Soit f une fonction 1-lipschitzienne de X dans $\mathbb{R}$. Alors

$$\mu\left(\{x \in X,\ |f(x) - \mathbb{E}_\mu f| \geqslant t\}\right) \leqslant 2 \exp -\frac{t^2}{2D^2}$$

où $D = 1/\sqrt{N}$ et où $\mathbb{E}_\mu f$ est la moyenne de f.

Ce théorème illustre l'idée que lorsqu'on a un espace de dimension N et de taille 1 (ici le diamètre du cube est 1), si chaque variable influe d'au plus $1/N$ alors les fluctuations du résultat sont d'ordre $1/\sqrt{N}$. La généralité de ce principe apparaît dans le(s) théorème(s) suivant(s) :

Théorème 23.
Le théorème 22 est valable pour les espaces métriques mesurés (X, d, μ) suivants, outre le cube :
- *la sphère S^N de dimension N et de rayon 1, avec pour μ la mesure de volume normalisée (Lévy [Lév22]) ;*
- *l'espace $\mathbb{R}^N$ muni de la mesure gaussienne $\mu(\mathrm{d}x) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-|x|^2/2\sigma^2\right)$ où l'écart-type σ est pris de telle sorte que $\mathbb{E}_\mu |x|^2 = 1$;*
- *toute variété riemannienne de dimension N dont la courbure de Ricci est au moins égale à celle de la sphère S^N, avec pour μ la mesure riemannienne normalisée (Gromov–Lévy [Gro86]).*

(La valeur optimale de l'écart-type D peut varier d'un petit facteur numérique selon les cas, mais est toujours du même ordre $\approx 1/\sqrt{N}$.)

Un théorème de concentration. Mais les différentes instances de ce théorème utilisaient pour leur démonstration des ingrédients variées (méthode de martingales, isopérimétrie sur la sphère ou l'espace gaussien, compréhension du rôle de la courbure de Ricci sur les variations de volume). Un de nos objectifs est de les faire apparaître comme des cas particuliers d'un même théorème utilisant la courbure de Ricci discrète.

Pour énoncer un tel théorème, nous aurons besoin d'introduire quelques quantités. Revenons à notre espace métrique (X, d) muni d'une famille de mesures $(m_x)_{x\in X}$, vue comme une marche aléatoire. La *constante de diffusion discrète* $\sigma(x)$ au point x est définie comme la distance quadratique moyenne entre deux points sous m_x :

$$\sigma(x) := \left(\frac{1}{2}\iint d(y,z)^2\, m_x(\mathrm{d}y)\, m_x(\mathrm{d}z)\right)^{1/2}$$

et la *dimension locale* en x est définie par

$$n_x := \frac{\iint d(y,z)^2\, m_x(\mathrm{d}y)\, m_x(\mathrm{d}z)}{\sup\left\{\iint d(f(y),f(z))^2\, m_x(\mathrm{d}y)\, m_x(\mathrm{d}z),\ f : X \to \mathbb{R}\ \ 1\text{−lipschitzienne}\right\}}$$

Par exemple, quand on prend pour m_x la marche aléatoire de pas ε dans une variété riemannienne de dimension N, on obtient que $\sigma(x)$ vaut environ ε et n_x vaut environ N (à des petites constantes numériques près). Définissons également

$$\sigma_\infty = \sup_{x\in X} \operatorname{diam} \operatorname{Supp} m_x$$

qui représente la *granularité* de notre marche aléatoire.

Nous pouvons maintenant énoncer un théorème affirmant que si un espace est à courbure de Ricci discrète positive, alors cet espace est concentré, regroupant tous les exemples précédents. Toutefois, il est facile de construire des exemples d'espaces de courbure positive où la concentration de la mesure est exponentielle et non gaussienne. Le théorème suivant permet de séparer un régime gaussien et un régime exponentiel.

Théorème 24 (Concentration de la mesure en courbure positive). *Soit (X, d) un espace métrique muni d'une marche aléatoire $(m_x)_{x \in X}$. Supposons que pour tous points $x, y \in X$ on ait $\kappa(x, y) \geqslant \kappa > 0$. La marche aléatoire (m_x) admet alors une unique mesure de probabilité invariante, que nous noterons μ.*

Soit

$$D_x^2 := \frac{\sigma(x)^2}{n_x \kappa}$$

et

$$D^2 := \mathbb{E}_\mu D_x^2$$

et supposons que la fonction $x \mapsto D_x^2$ est C-lipschitzienne et que $\sigma_\infty < \infty$. Posons

$$t_{\max} := \frac{D^2}{\max(\sigma_\infty, 2C/3)}$$

Alors, pour toute fonction 1-lipschitzienne $f : X \to \mathbb{R}$, pour $0 \leqslant t \leqslant t_{\max}$ on a

$$\mu\left(\{x, f(x) \geqslant \mathbb{E}_\mu f + t\}\right) \leqslant \exp -\frac{t^2}{6D^2}$$

et pour $t \geqslant t_{\max}$

$$\mu\left(\{x, f(x) \geqslant \mathbb{E}_\mu f + t\}\right) \leqslant \exp\left(-\frac{t_{\max}^2}{6D^2} - \frac{t - t_{\max}}{\max(3\sigma_\infty, 2C)}\right)$$

Exemples. Illustrons la manière dont ce théorème s'applique. Les constantes numériques de ce théorème ne sont pas optimales et offrent peu d'intérêt, et on les absorbera dans la notation $\approx$, qui signifie une égalité à une constante multiplicative près (constante ne dépendant d'aucun paramètre).

Soit m_x la marche aléatoire de pas ε sur une variété riemannienne X de dimension N à courbure de Ricci positive. Pour ε petit, la mesure μ est arbitrairement proche de la mesure de volume normalisée. On a $\sigma(x) \approx \varepsilon$ et $n_x \approx N$ pour tout x, et on a vu (exemple 16) que $\kappa \approx \frac{\varepsilon^2 \inf \mathrm{Ric}}{N}$ où $\inf \mathrm{Ric}$ est l'infimum de la courbure de Ricci ordinaire sur la variété. On obtient donc $D^2 \approx 1/\inf \mathrm{Ric}$ (indépendamment de ε assez petit) comme dans le théorème de Gromov–Lévy. Il n'y a pas de régime exponentiel : en effet on a $\sigma_\infty = 2\varepsilon$ et on peut prendre $C = 0$, et donc $t_{\max}$ tend vers l'infini pour ε petit, ce qui signifie que seul le régime gaussien apparaît.

Soit m_x la marche aléatoire de pas $1/N$ sur le cube discret $\{P, F\}^N$ (toujours muni de la métrique L^1 où chaque arête est de taille $1/N$). On a vu que $\kappa \approx 1/N$. Ici on a $\sigma(x) \approx 1/N$ et $n_x \approx 1$, et donc on obtient $D^2 \approx 1/N$, la même variance que dans le théorème de concentration sur le cube. De plus on a $\sigma_\infty \approx 1/N$ et $C = 0$, et donc

$t_{\max} \approx 1$ qui est du même ordre que le diamètre du cube. Le régime exponentiel n'est donc pas visible.

Notre dernier exemple illustre l'aspect exponentiel. Observons tout d'abord que le théorème passe très bien à la limite de marches aléatoires à temps continu. En effet, prenons un paramètre α très petit, et remplaçons m_x par $(1-\alpha)\delta_x + \alpha m_x$, qui est la même marche aléatoire avec un taux de saut α par unité de temps. Il est facile de voir que lorsque $\alpha \to 0$, $\sigma(x)$ se comporte comme $\sqrt{\alpha}$ tandis que κ se comporte comme α et n_x tend vers une constante, de sorte que D^2 a une limite finie.

Considérons alors une marche aléatoire à temps continu sur $\mathbb{N}$, dont la probabilité de transition de $k \in \mathbb{N}^*$ vers $k-1$ est $2k$ par unité de temps, tandis que la probabilité de transition de $k-1$ vers k est k par unité de temps. Il est facile de voir que la mesure géométrique $\mu(k) := 2^{-k+1}$ est invariante. On calcule immédiatement que $\sigma(k)^2$ vaut $3k+1$ par unité de temps, et que κ vaut 1 par unité de temps (et $n_k = 1$). On trouve alors $D_k^2 \approx k$ et $D^2 \approx 1$. Mais ici on a $C = 3$ et $\sigma_\infty = 2$ et donc $t_{\max} \approx 1$ de sorte que le régime gaussien n'existe pas.

Amusons-nous à tensoriser cet exemple N fois, c'est-à-dire à considérer la marche aléatoire produit sur $\mathbb{N}^N$ (muni de la métrique L^1) dont la projection sur chaque composante est celle ci-dessus. On constate alors que σ^2 est multiplié par N mais que κ ne change pas, et donc, la variance D^2 est multipliée par N comme on pouvait s'y attendre. Le point intéressant est que $t_{\max} \approx N$, ce qui veut dire qu'un régime gaussien apparaît quand N est grand. On obtient ainsi une version quantitative du théorème central limite, avec une estimation assez précise de la largeur de la fenêtre gaussienne avant le régime exponentiel (qui ne disparaît jamais totalement, puisque chaque composante reste de loi géométrique). Sur cet exemple, la transition entre les régimes gaussien et exponentiel peut en fait être calculée explicitement, et on vérifie que les ordres de grandeur sont les bons.

Commentaires. L'hypothèse que D_x^2 est une fonction lipschitzienne (ou, en fait, bornée par une fonction lipschitzienne) revient à dire que la constante de diffusion discrète croît au plus linéairement. C'est une hypothèse bien connue dans le monde des diffusions ou des marches aléatoires, puisque c'est celle sous laquelle sont en général énoncés les théorèmes d'existence d'un processus aléatoire étant donné son générateur. Sans les hypothèses que D_x^2 est lipschitzienne et que la granularité σ_∞ est finie, on peut facilement fabriquer des exemples de courbure positive où la concentration de la mesure n'est ni gaussienne ni exponentielle.

Un des points importants du résultat est que la variance D^2 obtenue est la *moyenne* des estimées locales D_x^2, et non un sup (le sup était infini dans notre dernier exemple). Ceci rappelle le théorème d'Efron–Stein (qui porte uniquement sur la variance, pas sur la concentration) bien connu des statisticiens [Mas07]. Dans le cas particulier des espaces produits, Lugosi et Massart [BLM03] ont démontré une version « gaussienne-exponentielle » du théorème d'Efron–Stein. Notre énoncé est donc dans le même esprit, mais ici la courbure de Ricci positive permet d'aller au-delà des espaces produits ; d'une certaine manière, la courbure positive permet de généraliser l'indépendance, et de traiter une sphère ou une gaussienne de la même manière.

Enfin, notons qu'une fois que les bonnes notions ont été posées, la démonstration du théorème 24 est relativement simple.

3.3 Autres résultats en courbure positive

Les espaces dont la courbure de Ricci discrète est minorée par une constante strictement positive satisfont d'autres propriétés, outre la concentration. Nous ne ferons ici que les survoler.

Dans un certain nombre de cas, il y a une petite perte dans les constantes numériques par rapport aux théorèmes riemanniens correspondants. Ceci doit être attendu dans un contexte qui ne distingue pas discret et continu ; les inégalités obtenues sont souvent optimales pour certains exemples discrets. En tout état de cause la perte sur les constantes est bornée (par un facteur 8).

– La définition a plusieurs conséquences immédiates. Notons l'existence d'une unique distribution invariante μ, vers laquelle la marche aléatoire donnée par les (m_x) converge exponentiellement vite en distance de transport. Cela permet de retrouver très rapidement des estimations de *temps de mélange* [DS96] pour les chaînes de Markov. On démontre aussi très simplement (dualité de Kantorovich [Vil03]) que la marche aléatoire est contractante en norme lipschitz, ou encore que la courbure de Ricci discrète satisfait une propriété de tensorisation très commode. Enfin, si la distance de transport entre x et m_x est uniformément bornée, l'espace est borné, ce qui est analogue au théorème de Bonnet–Myers pour les variétés (théorème 4).

– Comme la définition de la courbure de Ricci discrète ne fait intervenir que des inégalités portant sur la distance et les mesures, il est très facile de montrer que si une suite d'espaces métriques converge au sens de Gromov–Hausdorff [BBI01] vers un espace métrique donné, et si bien sûr les mesures (m_x) convergent également, alors la courbure de Ricci discrète converge aussi.

– Sous certaines hypothèses techniques, la valeur de la courbure de Ricci discrète donne une borne inférieure pour le *trou spectral* du laplacien naturellement associé à la marche aléatoire (m_x). Ceci généralise le *théorème de Lichnerowicz* pour les variétés riemanniennes.

– On peut généraliser les principaux résultats de la théorie de Bakry et Émery, à savoir la *contraction de gradient* et l'*inégalité de Sobolev logarithmique* (on renvoie à [ABCFGMRS00] pour une introduction à cette dernière). Pour cela, il faut utiliser une notion de « norme du gradient semi-locale » dépendant d'un paramètre $\lambda > 0$. On pose

$$\nabla_\lambda f(x) := \sup_{y,y' \in X} \frac{|f(y) - f(y')|}{d(y,y')} \, \mathrm{e}^{-\lambda d(x,y) - \lambda d(y,y')}$$

qui est en quelque sorte une constante de Lipschitz « autour de x » et redonne la norme du gradient ordinaire lorsque $\lambda \to \infty$. Mais selon la situation, il y a une valeur maximale de λ utilisable : pour une variété on peut prendre $\lambda \to \infty$ et récupérer le gradient habituel, mais pour un graphe on doit prendre λ d'ordre 1. Avec ce gradient, on peut alors démontrer l'inégalité de Sobolev logarithmique

$$\operatorname{Ent}_\mu f := \int f \log f \, \mathrm{d}\mu \leqslant \left(\sup_x \frac{4\sigma(x)^2}{\kappa n_x} \right) \int \frac{(\nabla_\lambda f)^2}{f} \, \mathrm{d}\mu$$

pour toute fonction $f : X \to \mathbb{R}_+^*$ telle que $\nabla_\lambda f < \infty$. (Plus exactement, il s'agit d'une inégalité de Sobolev logarithmique modifiée [BL98], puisque certains des exemples de courbure positive sont précisément des espaces pour lesquels les inégalités de Sobolev logarithmiques modifiées ont été introduites.) On démontre aussi l'inégalité de contraction de gradient

$$\nabla_\lambda(P^t f) \leqslant (1 - \tfrac{\kappa}{2})^t \, P^t(\nabla_\lambda f)$$

où P^t (pour t entier) est l'opérateur de moyenne associé à la marche aléatoire m_x.
À noter qu'en principe l'inégalité de Sobolev logarithmique implique la concentration gaussienne, par un argument bien connu de Herbst. Or on a vu qu'ici la concentration pouvait parfois être exponentielle plutôt que gaussienne. Mais l'utilisation du gradient semi-local ∇_λ fait que l'argument de Herbst s'interrompt justement à cette valeur de λ, et donne précisément de la concentration gaussienne puis exponentielle.

– La courbure nulle n'implique pas la concentration de la mesure, comme on le voit immédiatement pour l'espace euclidien. Cependant, la courbure positive ou nulle combinée à l'existence d'un point « localement attractif » pour la marche aléatoire implique la concentration exponentielle. L'exemple le plus élémentaire est ici la marche aléatoire sur $\mathbb{N}$ qui va vers la gauche avec probabilité $p > 1/2$ et vers la droite avec probabilité $1 - p$, pour laquelle le point 0 est localement attractif : la courbure est nulle partout sauf en 0 où elle est strictement positive, et la mesure invariante est géométrique.

– Dans un travail en cours avec Aldéric Joulin, étendant l'une de ses idées précédentes [Jou], nous utilisons la courbure discrète positive pour obtenir des bornes sur la *convergence des moyennes empiriques* pour les chaînes de Markov. Un point important est que ces bornes sont non asymptotiques et peuvent être évaluées en pratique à partir de quantités connues a priori en fonction de la chaîne de Markov.

3.4 Quelques problèmes ouverts

Mentionnons pour finir quelques questions qui se posent naturellement au vu de ces résultats. (La numérotation adoptée reprend celle de mon texte *A survey of Ricci* *curvature for metric spaces and Markov chains*, dont cette introduction est en partie une traduction.)

Problème A (Mesures log-concaves).
Nous avons vu que la courbure de Ricci de l'espace euclidien muni d'une mesure gaussienne est positive, et cela se généralise à toute mesure uniformément strictement log-concave assez lisse. Qu'en est-il pour une mesure log-concave quelconque, ou pour un ensemble convexe (muni par exemple du brownien réfléchi au bord), dont le bord est de courbure positive en un sens intuitif ?

Problème B (Variétés finsleriennes).
Nous avons vu que la courbure de Ricci discrète de $\mathbb{R}^N$ muni d'une distance L^p est nulle. Est-ce que, plus généralement, cette notion est intéressante dans les variétés finsleriennes, munies d'un processus pertinent ? (Voir par exemple [OS].)

Problème C (Groupes nilpotents).
Nous avons vu que la courbure de $\mathbb{Z}^N$ est nulle. Qu'en est-il pour des groupes nilpotents discrets ou continus ? Par exemple, sur le groupe de Heisenberg discret $\langle\, a, b, c \mid ac = ca, bc = cb, [a, b] = c\,\rangle$, la marche aléatoire naturelle correspondant au laplacien hypoelliptique du groupe de Heisenberg continu est engendrée par a et b. Comme la plus petite relation entre ces générateurs est de longueur 8, il est clair que la courbure de Ricci discrète est négative à petite échelle, mais tend-elle vers 0 aux plus grandes échelles ?

Problème T (Expanseurs).
Existe-t-il une famille de graphes expanseurs à courbure de Ricci discrète positive ou nulle ? (Une famille de graphes expanseurs est une famille de graphes finis, de taille tendant vers l'infini, de degré borné, et dont le trou spectral est uniformément minoré. Intuitivement, de tels graphes sont plutôt à courbure négative.)

Problème L (Théorème de Bishop–Gromov, etc.).
Est-il possible de généraliser davantage de résultats riemanniens traditionnels utilisant la courbure de Ricci ? Par exemple, peut-on obtenir un théorème de comparaison de volume analogue à Bishop–Gromov, et peut-on retrouver l'aspect isopérimétrique du théorème de Gromov–Lévy (dont nous n'avons généralisé que l'aspect concentration) ? Le problème est que ces théorèmes utilisent comme point de comparaison un espace de référence (la sphère d'une certaine dimension), ce qui est pertinent pour une variété mais sans doute pas pour un espace métrique discret comme le cube. Par exemple, dans le cube, la croissance de la taille des boules est exponentielle aux petits rayons, ce qui diffère fortement du comportement d'une variété.

Problème K (Définition de Sturm–Lott–Villani).
Y a-t-il un rapport quelconque entre notre notion de courbure de Ricci discrète et la notion de borne inférieure pour la courbure de Ricci définie par Sturm et Lott–Villani [Stu06, LV] ? Cette dernière semble s'appliquer surtout à des limites de variétés riemanniennes et est difficilement adaptable à des espaces discrets [BS], mais des résultats

supplémentaires ont été démontrés. Par exemple, le critère de courbure-dimension $\mathrm{CD}(K, n)$ implique l'inégalité de Brunn–Minkowski et le théorème de comparaison de Bishop–Gromov. Notre notion de courbure de Ricci discrète semble comparable à la condition plus faible $\mathrm{CD}(K, \infty)$.

Problème F (Constante optimale dans le théorème de Lichnerowicz).
Le théorème de Lichnerowicz minore le trou spectral du laplacien sur une variété riemannienne par $\frac{N}{N-1}$ fois l'infimum de la courbure de Ricci. Notre méthode appliquée à la marche aléatoire de pas ε sur une variété riemannienne donne seulement l'infimum de la courbure de Ricci, sans le facteur $\frac{N}{N-1}$. Ce résultat est optimal en toute généralité (par exemple pour le cube discret ou le processus d'Ornstein–Uhlenbeck, où on a égalité), mais peut-on exprimer dans notre langage une propriété qui permettrait d'obtenir le facteur $\frac{N}{N-1}$ dans le cas des variétés ?

Problème R (Dimension et théorème de Bonnet–Myers).
Le théorème qui nous sert d'analogue au théorème de Bonnet–Myers ressemble à une version L^1 de ce dernier plutôt qu'à une généralisation à proprement parler. Il est optimal en toute généralité (par exemple pour le cube). Nous avons donné une condition plus forte, inspirée de très près par le cas des variétés, qui permet d'obtenir un énoncé beaucoup plus proche du théorème de Bonnet–Myers classique, à savoir la condition

$$W_1(m_x^{*t}, m_{x'}^{*t'}) \leqslant \mathrm{e}^{-\kappa \inf(t,t')} d(x, x') + \frac{C(\sqrt{t} - \sqrt{t'})^2}{2d(x, x')}$$

pour toute paire de points x, x' et pour toute paire de temps t, t' assez petits (auparavant on ne considérait que $t = t'$). La constante C ressemble à une dimension. Est-elle reliée à la « dimension » dans la condition $\mathrm{CD}(K, n)$ de Bakry–Émery ?

Problème E (Trou spectral non réversible).
L'estimation du trou spectral par la courbure de Ricci discrète que nous avons obtenue n'est démontrée que lorsque la mesure invariante μ est *réversible* par rapport à la marche aléatoire (m_x), hypothèse couramment admise dans l'étude des marches aléatoires, ou bien lorsque X est fini. Cette hypothèse est-elle nécessaire ? (Dans le cas non réversible, il y a plusieurs manières a priori non équivalentes de définir le trou spectral.)

Problème M (Décroissance de l'entropie).
L'inégalité de Sobolev logarithmique non modifiée (i.e. celle comparant $\operatorname{Ent} f^2$ à $\int |\nabla f|^2$, et non $\operatorname{Ent} f$ à $\int |\nabla f|^2 / f$) entraîne classiquement une décroissance exponentielle de l'entropie sous l'action de la marche aléatoire. Qu'en est-il dans notre cadre ? Une fois de plus, il faut garder en tête certains exemples de courbure positive comme les distributions binomiales, pour lesquelles la forme modifiée de l'inégalité de Sobolev logarithmique avait dû être introduite.

Problème S (Espaces d'Alexandrov).
Les espaces de courbure sectionnelle positive au sens d'Alexandrov sont-ils de courbure de Ricci discrète positive, pour un certain choix de (m_x) ? Cela semble également une question ouverte avec la définition de Sturm–Lott–Villani.

Problème P (Courbure sectionnelle discrète).
Une notion analogue à la courbure sectionnelle positive serait de demander qu'il existe un transport de m_x vers m_y tel que *tous les points* sont déplacés d'*au plus* $d(x,y)$, et non seulement en moyenne. Cela revient à remplacer la distance W_1 par la distance de Wasserstein W_∞. Obtient-on des propriétés intéressantes ? Peut-on modifier cette définition pour obtenir une valeur non nulle pour cette courbure ? (Avec cette définition, la contribution de x et y dans m_x et m_y empêchera en général d'avoir une courbure strictement positive.) Y a-t-il un rapport avec la courbure sectionnelle positive au sens d'Alexandrov ?

Problème Q (Courbure scalaire discrète).
En géométrie riemannienne, la courbure scalaire discrète, dont nous n'avions pas encore parlé, est la moyenne de la courbure de Ricci sur tous les vecteurs tangents en un point. Elle contrôle par exemple la croissance du volume des boules. Ici on pourrait poser $S(x) := \int \kappa(x,y)\, m_x(\mathrm{d}y)$. Que peut-on en dire ?

Problème N (Flot de Ricci discret).
On peut tenter de définir un « flot de Ricci discret » en changeant la distance sur X au cours du temps en fonction de la courbure :

$$\frac{\mathrm{d}}{\mathrm{d}t} d(x,y) = -\kappa(x,y)\, d(x,y)$$

où $\kappa(x,y)$ est calculé en utilisant la valeur courante de la distance $d(x,y)$. Cela revient à remplacer la distance $d(x,y)$ par $(1-\mathrm{d}t)\, d(x,y) + \mathrm{d}t\; W_1(m_x, m_y)$ (voir aussi [MT]). Il n'est pas évident si l'on doit conserver le même noyau de transition (m_x) au cours du temps, ou bien le faire évoluer en fonction de la distance (ce qui correspond plus au flot de Ricci ordinaire). Peut-on en dire quelque chose d'intéressant ?

Problème O (À δ près).
L'inégalité $W_1(x,y) \leqslant (1-\kappa)\, d(x,y)$ est contraignante lorsque x et y sont très proches. Pour éliminer l'influence des très petites échelles, on peut s'inspirer de la définition des espaces δ-hyperboliques et définir une « courbure de Ricci positive à δ près » par l'inégalité

$$W_1(x,y) \leqslant (1-\kappa)\, d(x,y) + \delta$$

Alors la courbure positive devient une propriété *ouverte* en topologie de Gromov–Hausdorff. Lesquels de nos résultats s'étendent à ce cadre ?

4 Courbure à grande échelle : physique statistique relativiste

Par physique statistique relativiste, on entend en général l'étude des objets classiques de la physique statistique (fluides ou autres systèmes de particules) dans un contexte où la relativité ne peut pas être négligée. Mais ici, c'est sur l'espace-temps lui-même que nous ferons la statistique. Nous considérerons donc les effets sur l'espace-temps de diverses fluctuations, comme par exemple des ondes gravitationnelles ou des fluctuations de la densité de matière.

L'objectif est d'obtenir une description effective de l'influence moyenne de ces fluctuations à des échelles où elles sont trop petites pour être directement observables (comme l'échelle de l'univers lui-même). En effet, la théorie de la relativité générale étant non linéaire, il n'est pas vrai que des fluctuations de moyenne nulle produisent un effet moyen nul sur la dynamique à grande échelle de l'univers, comme cela serait le cas avec la théorie de Newton.

L'équation d'Einstein liant la courbure de Ricci au contenu en matière de l'espace-temps, on aura besoin de comparer la courbure de Ricci d'une métrique régulière à celle de la même métrique perturbée. La non-linéarité du passage de la métrique vers la courbure produira un effet a priori non nul en moyenne, qui s'interprète par l'équation d'Einstein comme une *matière apparente* qui résume les effets moyens des irrégularités.

Il s'agit non d'une étude mathématique, mais bien d'isoler un effet physique. En particulier, nous ne présenterons pas de théorèmes. L'approche utilisée ici a été initiée par Fabrice Debbasch [Deb04, Deb05] (après d'autres tentatives, voir par exemple [Zal97, Buc00]), qui a défini en toute généralité un modèle de champ moyen pour la relativité générale.

4.1 Un peu de relativité générale

Commençons par quelques rappels trop rapides de relativité générale. Nous renvoyons par exemple à l'ouvrage de référence [Wal84].

Notation d'Einstein pour le calcul tensoriel. On adoptera ici la *notation d'Einstein* pour noter les vecteurs tangents et les tenseurs sur une variété. Cette notation consiste à faire suivre chaque objet d'indices arbitraires indiquant son type ainsi que les différentes contractions entre l'espace et son dual.

Soit E un espace vectoriel de dimension finie (par exemple, l'espace tangent à une variété M en un point x). On notera un élément de E avec un indice (symbole arbitraire souvent pris dans l'alphabet grec) en haut. On notera une forme linéaire sur E, c'est-à-dire un élément de E^*, avec un indice arbitraire en bas. Une forme bilinéaire sur E est un élément de $E^* \otimes E^*$ et portera ainsi deux indices en bas.

Si on dispose d'un vecteur $v^\alpha \in E$ et d'une forme linéaire $f_\beta \in E^*$, on peut bien sûr évaluer la forme sur le vecteur, ce qui est une application $E^* \otimes E \to \mathbb{R}$. Cette opération est indiquée par la répétition d'indice, ainsi, $f_\alpha v^\alpha$ est un élément de $\mathbb{R}$. Par contre, $f_\alpha v^\beta$ est simplement un élément de $E^* \otimes E$ égal au produit tensoriel de f_α et v^β.

Notons que (en dimension finie) l'espace $\mathrm{L}(E)$ des applications linéaires de E dans E s'identifie à $E \otimes E^*$. Ainsi, si $f_\mu^\nu \in E \otimes E^*$ est une application linéaire et $v^\alpha \in E$ un vecteur, l'image de v^α par f_μ^ν est le vecteur $f_\alpha^\nu v^\alpha \in E$ obtenu en évaluant la deuxième composante de f_μ^ν sur le vecteur v^α. (Remarquer bien sûr la similarité avec l'application d'une matrice à un vecteur.) L'identité de E dans E est en général notée δ_μ^ν.

La *trace* est simplement l'application de $\mathrm{L}(E) \simeq E \otimes E^*$ dans $\mathbb{R}$ qui consiste à évaluer la deuxième composante sur la première. Ainsi, la trace de $f_\mu^\nu \in E \otimes E^*$ est f_μ^μ.

C'est alors un simple exercice de vérifier qu'étant donné une base de E, on peut trouver les coordonnées d'un tenseur dans cette base en interprétant chaque indice comme un numéro dans la base, et en sommant sur les indices répétés.

Par exemple, soit $g_{\mu\nu} \in E^* \otimes E^*$ une forme bilinéaire sur E. Alors si v^α, w^β sont des vecteurs de E, l'objet $g_{\mu\nu} v^\alpha w^\beta$ appartient à $E^* \otimes E^* \otimes E \otimes E$, tandis que le nombre $g_{\mu\nu} v^\mu w^\nu$ est l'image du précédent par l'évaluation de la première composante sur la troisième et de la deuxième sur la quatrième, qui est tout simplement le produit scalaire de v^μ et w^ν pour la forme quadratique $g_{\mu\nu}$. L'objet $g_{\mu\nu} v^\mu \in E^*$ est, lui, une forme linéaire qui, évaluée sur un vecteur, renverra son produit scalaire avec v^μ. Si, en outre, la forme bilinéaire $g_{\mu\nu} \in E^* \otimes E^* \simeq \mathrm{L}(E, E^*)$ est non dégénérée, elle a un inverse qui est un élément de $\mathrm{L}(E^*, E) \simeq E \otimes E$, noté généralement $g^{\mu\nu}$ avec le même nom mais les indices en haut.

Ces notations seront utilisées par la suite dans le cas où E est l'espace tangent à une variété en un point donné. Les éléments de $E^{\otimes p} \otimes (E^*)^{\otimes q}$ seront appelés (p, q)-tenseurs. Par exemple, une métrique sur une variété est un une forme bilinéaire c'est-à-dire un $(0, 2)$-tenseur.

Cette notation est très utile en géométrie riemannienne ou lorentzienne. (Par exemple, si ∇_α est la connexion de Levi-Civita, alors $\nabla_\alpha f$ est la différentielle de la fonction f, sa hessienne est la forme bilinéaire $\nabla_\alpha \nabla_\beta f$, son laplacien est $g^{\alpha\beta} \nabla_\alpha \nabla_\beta f$; la divergence d'un champ de vecteurs v^β est $\nabla_\beta v^\beta$, et l'opérateur de courbure de Riemann est juste $\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha$ sans qu'il y ait besoin de crochets de Lie dans la définition...)

Variétés lorentziennes. Rappelons qu'une variété munie d'une *métrique lorentzienne* est une variété de dimension $N + 1$ munie, en chaque point, d'une forme bilinéaire de signature $(N, 1)$, de sorte que localement on peut trouver des coordonnées $(t, x_1, \ldots, x_N)$ où la métrique est de la forme $-\mathrm{d}t^2 + \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_N^2$. (On a pris la vitesse de la lumière égale à 1.) Nous ne discuterons pas ici les raisons physiques de ce choix, qui sont décrites dans tout texte d'introduction à la relativité restreinte.

Nous avons défini plus haut la courbure de Ricci d'une variété riemannienne. La définition que nous avons donnée en mesurant des distances ne se transpose pas directement au cas lorentzien (une variété lorentzienne n'est pas un espace métrique), mais les formules explicites qu'on obtient pour calculer la courbure dans le cas riemannien ont encore un sens dans le cas lorentzien, et ce c'est par elles que l'on définit la courbure de Ricci.

Nous avons vu que la courbure de Ricci le long d'un vecteur tangent v est une forme quadratique $\mathrm{Ric}(v, v)$. La courbure de Ricci définit ainsi un $(0, 2)$-tenseur, noté le plus souvent $R_{\mu\nu}$, tel que la courbure de Ricci le long de v^μ est $R_{\mu\nu}v^\mu v^\nu$.

L'équation d'Einstein. Un *espace-temps* est la donnée d'une variété M de dimension 4, d'une métrique lorentzienne $g_{\mu\nu}$ sur M, et d'un $(0, 2)$-tenseur T appelé *tenseur d'énergie-impulsion* de la matière, tel que l'*équation d'Einstein* soit satisfaite

$$R_{\mu\nu} - \frac{1}{2} R_{\alpha\beta}\, g^{\alpha\beta}\, g_{\mu\nu} = 8\pi\chi T_{\mu\nu}$$

où $R_{\mu\nu}$ est le tenseur de Ricci de la métrique $g_{\mu\nu}$ et χ est une constante reliée à la constante de gravitation ($\chi = 1$ dans les unités canoniques).

Bien sûr, on peut toujours se donner une métrique $g_{\mu\nu}$ arbitraire, calculer le tenseur d'Einstein $R_{\mu\nu} - \frac{1}{2} R_{\alpha\beta} g^{\alpha\beta} g_{\mu\nu}$ associé à $g_{\mu\nu}$, et choisir $T_{\mu\nu}$ en fonction, mais on souhaite que le tenseur d'énergie-impulsion obtenu soit physiquement pertinent et décrive la matière.

Dans le cas classique, le tenseur d'énergie-impulsion d'un fluide de densité ρ et de vitesse v est $\rho\, v \otimes v$, et décrit les flux de la quantité de mouvement ρv. En relativité générale on cherchera donc très souvent des tenseurs d'énergie-impulsion de la forme $T_{\mu\nu} = \rho u_\mu u_\nu$ où u_μ est la quadrivitesse du fluide. Une forme un peu plus générale prenant en compte la pression P est

$$T_{\mu\nu} = \rho u_\mu u_\nu + P\,(g_{\mu\nu} + u_\mu u_\nu)$$

(noter que $g_{\mu\nu} + u_\mu u_\nu$ est simplement la métrique restreinte à l'orthogonal de u_μ, autrement dit la partie spatiale de la métrique dans le référentiel de la particule).

Par exemple, dans les modèles les plus simples, à l'échelle de l'univers les galaxies sont traitées comme un fluide de pression nulle (« poussière »).

On peut aussi décrire le tenseur d'énergie-impulsion associé à un champ électromagnétique ou d'autres types de champs, qui ne seront pas mentionnés ici.

Pour différentes raisons (physiquement, des raisons d'homogénéité, ou mathématiquement parce que l'impulsion est plutôt un élément de l'espace cotangent et donc le tenseur d'énergie-impulsion est $p \otimes v$), on préfère multiplier les objets intervenant dans l'équation d'Einstein par l'inverse de la métrique et écrire cette équation sous la forme

$$R_\mu^\nu - \frac{1}{2} R_\alpha^\alpha\, \delta_\mu^\nu = 8\pi\, T_\mu^\nu$$

En effet, les coordonnées du tenseur T_μ^ν dans un référentiel s'interprètent plus directement comme des quantités physiques.

4.2 Un modèle de champ moyen pour la relativité générale

Le modèle de champ moyen introduit dans [Deb04] consiste à se donner une variété M fixée, munie d'une métrique aléatoire. Par exemple, la métrique peut être celle d'un univers homogène et isotrope auquel on ajoute des ondes gravitationnelles de direction aléatoire. On suppose que l'observateur ne voit pas directement les fluctuations. Il faut alors comparer l'évolution moyenne des métriques ayant des ondes gravitationnelles, à celle de la métrique homogène de référence perçue par l'observateur.

Le modèle est le suivant. Soit M une variété fixée, et soit $g_{\mu\nu}$ une métrique aléatoire sur M ; la loi de $g_{\mu\nu}$ caractérise le type de fluctuations considérées. À la métrique aléatoire $g_{\mu\nu}$ correspond un tenseur d'énergie-impulsion aléatoire $T_{\mu\nu}$ par l'équation d'Einstein. En général, on définit plutôt la loi de $g_{\mu\nu}$ en se donnant un modèle pour les fluctuations de matière $T_{\mu\nu}$ et en essayant de définir $g_{\mu\nu}$ en fonction.

Définissons la métrique

$$\bar{g}_{\mu\nu} = \mathbb{E} g_{\mu\nu}$$

qui est une métrique sur M (en effet, l'espace des formes bilinéaires sur une variété donnée est un espace vectoriel). Pour des modèles de fluctuations raisonnables, $\bar{g}_{\mu\nu}$ est encore de signature $(3,1)$. L'idée est que $\bar{g}_{\mu\nu}$ représente la métrique moyenne perçue par l'observateur.

Cette métrique $\bar{g}_{\mu\nu}$ définit une courbure de Ricci $\bar{R}_\mu^\nu$, qui à son tour définit un tenseur d'énergie-impulsion $\bar{T}_\mu^\nu$ par l'équation d'Einstein

$$\bar{R}_\mu^\nu - \frac{1}{2}\bar{R}_\alpha^\alpha \delta_\mu^\nu = 8\pi\,\bar{T}_\mu^\nu$$

mais, comme la courbure de Ricci n'est pas une fonction linéaire de la métrique, on a a priori

$$\bar{T}_\mu^\nu \neq \mathbb{E} T_\mu^\nu$$

Autrement dit, un observateur ayant accès uniquement à la métrique moyenne $\bar{g}_{\mu\nu}$ et au contenu moyen en matière constatera une violation de l'équation d'Einstein.

Un cas amusant est celui où les fluctuations sont des ondes gravitationnelles : on a alors toujours $T_\mu^\nu = 0$, mais a priori $\bar{T}_\mu^\nu$ est non nul. En particulier, un univers sans matière mais avec des ondes gravitationnelles aléatoires peut se comporter, en moyenne, comme un univers contenant de la matière. Physiquement, l'idée est que l'énergie d'auto-interaction des ondes gravitationnelles avec elles-mêmes va en moyenne influer sur la dynamique à grande échelle de l'univers, comme si celui-ci contenait davantage de matière. Ceci évoque évidemment le problème de la matière noire.

On définit donc l'effet de « matière apparente » comme le terme supplémentaire que l'observateur doit ajouter pour que l'équation d'Einstein soit satisfaite :

$$T^{\mathrm{app}}{}_\mu^\nu := \bar{T}_\mu^\nu - \mathbb{E} T_\mu^\nu$$

et notre objectif est de quantifier cet effet.

4.3 La matière apparente dans l'univers

Avec F. Debbasch et C. Chevalier, nous avons estimé l'effet de matière apparente dans quelques situations. Par exemple, on peut faire un développement perturbatif autour de la métrique suivante, qui décrit un univers homogène et isotrope (modèle de Friedmann–Lemaître–Robertson–Walker) asymptotiquement plat, c'est-à-dire où la densité de matière est égale à la densité critique. La métrique est donnée par

$$\mathrm{d}s^2 = \eta^4 \left(-\mathrm{d}\eta^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right)$$

Le temps physique est défini par $\mathrm{d}t^2 = \eta^4 \mathrm{d}\eta^2$ soit $t = \eta^3/3$. Par l'équation d'Einstein on obtient le tenseur d'énergie-impulsion associé, et on vérifie qu'il est de la forme $\rho u_\mu u^\nu$ où la quadri-vitesse u_μ est $(\eta^2, 0, 0, 0)$ (la matière est statique dans ces coordonnées) et la densité $\rho = \frac{3}{2\pi\,\eta^6}$ dans ces coordonnées ; la pression est nulle.

Nous avons considéré des petites perturbations autour de cette métrique, représentant soit des ondes gravitationnelles, soit des variations de la densité et de la vitesse de la matière.

Les perturbations à l'ordre 1 étant linéaires par définition, elles commutent avec l'espérance $\mathbb{E}$, et l'effet de matière apparente est donc d'ordre 2. Cela correspond à l'intuition physique que ce phénomène est relativiste et non newtonien, et résulte de l'interaction du champ gravitationnel avec lui-même.

Les conclusions sont les suivantes : l'effet est beaucoup plus important pour des ondes gravitationnelles que pour des fluctuations de matière. Pour des ondes gravitationnelles, la matière apparente obtenue a un tenseur d'énergie-impulsion ${T^{\mathrm{app}}}_\mu^\nu$ qui se comporte comme celui d'un fluide dont la densité d'énergie est $\varepsilon^2\, n_{\mathrm{osc}}^2\, \rho/48$ et la pression $\varepsilon^2\, n_{\mathrm{osc}}^2\, \rho/144$, où ε est l'amplitude relative de l'onde gravitationnelle et n_{osc} le nombre de périodes de l'onde dans l'univers observable à un instant donné. Notons qu'à cet ordre, la relation entre énergie et pression est celle attendue pour de la *radiation* classique.

Quant aux fluctuations de la matière (vitesse, densité), elles engendrent une matière apparente se comportant à cet ordre comme le carré de l'amplitude relative de la perturbation, sans effet de fréquence. Selon les caractéristiques de la fluctuation, la matière apparente peut être d'énergie positive ou négative, et il en est de même pour sa pression.

En pratique, dans ce modèle une seule onde gravitationnelle d'amplitude relative $\approx 10^{-5}$ et de fréquence $\approx 10^{-12}\,\mathrm{Hz}$ (ce qui signifie que cette onde déforme un cercle parfait en une ellipse d'excentricité 10^{-5} au bout de cent mille ans) pourrait beaucoup changer la valeur effective de la densité d'énergie ρ. Une telle onde gravitationnelle n'est pas détectable avec les moyens techniques d'aujourd'hui [ZZ06].

Pour pouvoir donner des estimations plus précises, on doit bien sûr encore affiner ces calculs (qui sont perturbatifs) et en étendre la portée à des modèles d'univers un peu plus réalistes. Mais il semble clair que ces effets de matière apparente ne devraient pas être négligés en cosmologie.

*
* *

Nous espérons que ce parcours dans le monde des courbures et de quelques-unes de leurs applications aura distrait le lecteur. Peut-être même souhaitera-t-il étudier l'une des questions ouvertes que nous avons mentionnées...

Références

[ABCFGMRS00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses **10**, Société Mathématique de France (2000).

[BBI01] D. Burago, Y. Burago, S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics **33**, AMS (2001).

[BD97] R. Bubley, M. E. Dyer, *Path coupling: a technique for proving rapid mixing in Markov chains*, FOCS 1997, 223–231.

[BE84] D. Bakry, M. Émery, *Hypercontractivité de semi-groupes de diffusion*, C. R. Acad. Sci. Paris Sér. I Math. **299** (1984), n° 15, 775–778.

[BE85] D. Bakry, M. Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84. Lecture Notes in Math. **1123**, Springer, Berlin (1985), 177–206.

[Ber03] M. Berger, *A panoramic view of Riemannian geometry*, Springer, Berlin (2003).

[BH99] M. R. Bridson, A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der mathematischen Wissenschaften **319**, Springer (1999).

[BL98] S. Bobkov, M. Ledoux, *On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures*, J. Funct. Anal. **156** (1998), n° 2, 347–365.

[BLM03] S. Boucheron, G. Lugosi, P. Massart, *Concentration inequalities using the entropy method*, Ann. Probab. **31** (2003), n° 3, 1583–1614.

[BS] A.-I. Bonciocat, K.-T. Sturm, *Mass transportation and rough curvature bounds for discrete spaces*, preprint.

[Buc00] T. Buchert, *On average properties of inhomogeneous cosmologies*, Gen. Rel. Grav. **32** (2000), 105–126.

[Car92] M. P. do Carmo, *Riemannian geometry*, Mathematics: Theory & Applications, Birkhäuser (1992).

[Deb04] F. Debbasch, *What is a mean gravitational field?*, Eur. Phys. J. B **37** (2004), n° 2, 257–270.

[Deb05] F. Debbasch, *Mean field theory and geodesics in general relativity*, Eur. Phys. J. B **43** (2005), n° 1, 143–154.

[DGW04] H. Djellout, A. Guillin, L. Wu, *Transportation cost–information inequalities and applications to random dynamical systems and diffusions*, Ann. Prob. **32** (2004), n° 3B, 2702–2732.

[Dob70] R. L. Dobrušin, *Definition of a system of random variables by means of conditional distributions* (Russian), Teor. Verojatnost. i Primenen. **15** (1970), 469–497. English translation: *Prescribing a system of random variables by conditional expectations*, Theory of Probability and its Applications **15** (1970) n° 3, 458–486.

[Dob96] R. Dobrushin, *Perturbation methods of the theory of Gibbsian fields*, in R. Dobrushin, P. Groeneboom, M. Ledoux, *Lectures on probability theory and statistics*, Lectures from the 24th Saint-Flour Summer School held July 7–23, 1994, edited by P. Bernard, Lecture Notes in Mathematics **1648**, Springer, Berlin (1996), 1–66.

[DS85] R. L. Dobrushin, S. B. Shlosman, *Constructive criterion for the uniqueness of Gibbs field*, in J. Fritz, A. Jaffe and D. Szász (eds), *Statistical physics and dynamical systems*, papers from the second colloquium and workshop on random fields: rigorous results in statistical mechanics, held in Köszeg, August 26–September 1, 1984, Progress in Physics **10**, Birkhäuser, Boston (1985), 347–370.

[DS96] P. Diaconis, L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. **6** (1996), n° 3, 695–750.

[GhH90] É. Ghys, P. de la Harpe, *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math. **83**, Birkhäuser (1990).

[Ghy03] É. Ghys, *Groupes aléatoires*, séminaire Bourbaki **916** (2003). Paru dans Astérisque **294** (2004), 173–204.

[GrH97] R.I. Grigorchuk, P. de la Harpe, *On problems related to growth, entropy, and spectrum in group theory*, Dynam. Control Systems **3** (1997), n° 1, 51–89.

[Gro86] M. Gromov, *Isoperimetric inequalities in Riemannian manifolds*, in V. Milman, G. Schechtman, *Asymptotic theory of finite dimensional normed spaces*, Lecture Notes in Mathematics **1200**, Springer, Berlin (1986), 114–129.

[Gro87] M. Gromov, *Hyperbolic groups*, in *Essays in group theory*, ed. S. M. Gersten, Springer (1987), 75–265.

[Gro93] M. Gromov, *Asymptotic Invariants of Infinite Groups*, in *Geometric group theory*, ed. G. Niblo, M. Roller, Cambridge University Press, Cambridge (1993).

[Gro99] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*, Progress in Math. **152**, Birkhäuser (1999).

[Gro03] M. Gromov, *Random walk in random groups*, GAFA, Geom. Funct. Anal. **13** (2003), n°1, 73–146.

[Jou07] A. Joulin, *Poisson-type deviation inequalities for curved continuous time Markov chains*, Bernoulli **13** (2007), n°3, 782–798.

[Jou] A. Joulin, *A new Poisson-type deviation inequality for Markov jump process with positive Wasserstein curvature*, preprint.

[Led01] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs **89**, AMS (2001).

[Lév22] P. Lévy, *Leçons d'analyse fonctionnelle*, Gauthier-Villars, Paris (1922). Nouvelle édition en 1951 sous le titre *Problèmes concrets d'analyse fonctionnelle*.

[LV] J. Lott, C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, preprint.

[Mas07] P. Massart, *Concentration inequalities and model selection*, Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, Lecture Notes in Mathematics **1896**, Springer, 2007.

[MT] R. J. McCann, P. M. Topping, *Ricci flow, entropy and optimal transportation*, preprint.

[Oht07] S.-i. Ohta, *On the measure contraction property of metric measure spaces*, Comment. Math. Helv. **82** (2007), n° 4, 805–828.

[Oli] R. Imbuzeiro Oliveira, *On the convergence to equilibrium of Kac's random walk on matrices*, preprint, `arXiv:0705.2253`

[OS] S.-i. Ohta, K.-T. Sturm, *Heat flow on Finsler manifolds*, preprint.

[OV00] F. Otto, C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), 361–400.

[RS05] M.-K. von Renesse, K.-T. Sturm, *Transport inequalities, gradient estimates, and Ricci curvature*, Comm. Pure Appl. Math. **68** (2005), 923–940.

[Sch01] G. Schechtman, *Concentration, results and applications*, in *Handbook of the geometry of Banach spaces*, vol. 2, North-Holland, Amsterdam (2003), 1603–1634.

[Stu06] K.-T. Sturm, *On the geometry of metric measure spaces*, Acta Math. **196** (2006), n° 1, 65–177.

[Vil03] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics **58**, American Mathematical Society, Providence (2003).

[Wal84] R. M. Wald, *General relativity*, University of Chicago Press, Chicago (1984).

[Zal97] R. Zalaletdinov, *Averaging problem in general relativity, macroscopic gravity and using Einstein's equation in cosmology*, Bull. Astron. Soc. India **25** (1997), 401–416.

[ZZ06] W. Zhao, Y. Zhang, *Relic gravitational waves and their detection*, Phys. Rev. D **74** (2006), n° 4, 043503.

Dossier scientifique

Habilitation à diriger des recherches

A January 2005 invitation to random groups

Sans préjudice du titre, il a fallu une bonne partie de l'année 2005 pour écrire ce texte de synthèse, dont l'idée remonte à un séjour à Neuchâtel en janvier 2005, à l'invitation d'Alain Valette. Le tout forme un court livre publié, sur le conseil d'Étienne Ghys, comme volume 10 (2005) de la série Ensaios Matemáticos, *éditée par la Sociedade Brasileira de Matemática.*

Yann Ollivier, École normale supérieure de Lyon, 2009

A January 2005 invitation to random groups

Yann Ollivier

Abstract

Random groups provide a rigorous way to tackle such questions as "What does a typical (finitely generated) group look like?" or "What is the behavior of an element of a group when nothing particular happens?"

We review the results obtained on random groups as of January 2005. We give proper definitions and list known properties of typical groups. We also emphasize properties of random elements in a given group. In addition we present more specific, randomly twisted group constructions providing new, "wild" examples of groups.

A comprehensive discussion of open problems and perspectives is included.

Foreword

Our aim here is to present, within the limited scope of the author's knowledge, the state of the art of random groups. The decision to write such a survey arose from consideration of the rapidly growing number of publications on the subject, which, from a bunch of theorems, is slowly shaping into a theory. A whole section has been devoted to the statement of open problems of various difficulty.

The accompanying *Primer to geometric group theory*[1] is meant as a gentle introduction to the necessary background material, may readers outside of the field find some appeal in random groups.

There are no proofs in this book, except that of the foundational density 1/2 phase transition theorem, which constitutes a standalone chapter at the end of the text. Most results are indeed very technical and gathering all proofs would have resulted in a heavy (in all senses of the word) treatise rather than an "invitation".

The goal was not to track down the origins of the generic way of thinking in group theory or neighboring fields, but to review the results in that branch of mathematics which treats of the groups obtained from random presentations. In particular, and mainly because of the author's incompetence on these matters, the asymptotic theory of finite groups and properties of random elements therein are not covered.

The information presented here has deliberately been limited to the works available to the author as of January 31st, 2005, except for bibliographical references to then unpublished manuscripts, which have been updated for the reader's convenience.

[1] Partially reproduced in French in the first part of this Habilitation document

The roots of all current mathematical work related to random groups lie unquestionably in Misha Gromov's fertile mind, and can be traced back to his seminal 1987 paper [Gro87] on hyperbolic groups. In order to illustrate the importance of his newly defined [Gro78, Gro83] class of groups, he stated (without proof) that "most" groups with a fixed number of generators and relations and "long enough" relation length are hyperbolic (see § I.1.).

He later substantiated his thoughts on the subject in Chapter 9 of [Gro93], entitled *Finitely presented groups: density of random groups and other speculations*, where the *density model* of random groups is defined and the intuition behind it thoroughly discussed. This model allows a sharp control of the quantity of relations put in a random group, and has proven very fruitful over the years, especially since the properties obtained vary with density (cf. § I.2.).

The subject received considerable attention from the general mathematical community (see e.g. [Ghy03, Pan03]) when Gromov published his *Random walk in random groups* [Gro03] (elaborating on the equally renowned *Spaces and questions* [Gro00]), in which he uses random methods to build a group with "wild" geometric properties linked to the Baum-Connes conjecture with coefficients (see § III.2.), although these partially random groups have no pretention at all to model a "typical" group.

The first motivation for the study of random groups is the following somewhat philosophical question: "What does a typical group look like?" This theme is addressed in § I., **Models of typical groups**, where known properties of those are discussed. The word "typical" here is used as a convenient loose term interpolating between "random", which entails a probabilistic setting, and "generic", rather implying a topological framework. The latter is specifically developed in § I.4., where some results on the space of all marked groups are presented.

A slightly different approach is to look at properties of "typical" elements in a given group, either for themselves or in order to achieve certain goals. This is the theme of § II., **Typical elements in a group**. For example, a lot of "unrelated" "typical" elements in a hyperbolic group can be killed without harming too much the group (§ II.1.); this intuition has been present since the very beginning of hyperbolic group theory. Also, considering that typical relations in a presentation do not exhibit any special structure led to a sharp evaluation of the number of different one-relator groups (§ II.3.).

But random groups now have found applications to other fields of mathematics. Indeed, the use of random ingredients in constructions specifically designed to achieve certain goals allows to prove existence of groups with new properties, which are counterexamples to open questions, such as Gromov's celebrated group (§ III.2.) whose Cayley graph admits no uniform embedding into the Hilbert space, or a bunch of new groups with property (T) and somewhat unexpected properties (§ III.3.). Though these groups cannot pretend to be good candidates for "typicality", they are definitely of interest to people in and outside of group theory.

Yet for the author, the primary appeal of the field is still the study of properties of “typical” groups for themselves, rather than the applications just discussed. This is, of course, a matter of (philosophical?) taste.

Acknowledgements. I would like to thank Goulnara Arzhantseva, Vincent Beffara, Yves de Cornulier, Thomas Delzant, Christophe Deroulers, Étienne Ghys, Misha Gromov, Alice Guionnet, Frédéric Haglund, Ilya Kapovich, Richard Kenyon, Marc Mézard, Pierre Pansu, Pierre Senellart, Bruno Sévennec, Jean-Claude Sikorav, Alain Valette, Maria Eulália Vares and Dani Wise for their help and comments with preparing this book.

Contents

Notation and conventions

$\mathbb{N}$: the set of natural numbers, including 0.

$\#A$: number of elements of the set A.

F_m: free group of rank m over the set of generators $a_1, \ldots, a_m$ and their formal inverses. Unless otherwise stated, we assume $m \geqslant 2$.

$\langle R \rangle$: normal subgroup generated by the set of elements R.

$\langle\, a_1, \ldots, a_m \mid R\, \rangle$: group presented by generators $a_1, \ldots, a_m$ with set of relators R, that is, the group $F_m / \langle R \rangle$.

$|w|$: length of the word w.

$|D|$: number of faces of the van Kampen diagram D.

$|\partial D|$: boundary length of the van Kampen diagram D.

Reduced word: a word not containing any letter immediately followed by its inverse.

Non-elementary hyperbolic group: a hyperbolic group which is neither finite nor quasi-isometric to $\mathbb{Z}$.

Hyperbolicity: the fact of being non-elementary hyperbolic.

With overwhelming probability: with probability tending to 1 when some natural parameter (often denoted ℓ) tends to infinity.

Part I.
Models of typical groups

The basic idea of random groups is to take a group presentation at random and to look at what are "typically" the properties of the group so obtained, leading to such statements as "almost every group is hyperbolic". Of course this makes sense only if some precise way to pick presentations at random is prescribed: this is what we call a *model* of random groups.

So a random group will usually be given by a presentation by generators and relators

$$G = \langle\, S \mid R \,\rangle$$

where $S = \{a_1, \ldots, a_m\}$ is some finite[2] set of generators, and R is a set of words on the elements of S (and their inverses), taken at random. Since any group presentation can be written using reduced words only (i.e. words not containing $a_i a_i^{-1}$ or $a_i^{-1} a_i$), usually only such words are considered.

To choose a model of random groups is to specify a probability law for the set of relators R. Probability and statistics are most relevant when some parameter is large so that laws of large numbers can be used. In most (but not all) models, the set of generators $S = \{a_1, \ldots, a_m\}$ is kept fixed, and the large parameter is the length of the words in R. One more degree of freedom is to let the number of words in R grow as their length becomes larger. These choices allow for different models.

The models. Basically there are three models of random groups. The quite straightforward *few-relator model* (Def. 1) allows for only a fixed number of relators, of bounded length; small cancellation, hence hyperbolicity, is easily shown to be generic in this model. It is now subsumed as density 0 in the density model. The *few-relator model with various lengths* (Def. 4), which allows very different relator lengths, is more difficult technically because several scales are involved. The *density model* (Def. 7) allows a clear-cut quantitative approach on the number of relators that can be put before the group collapses; this model has been preferentially focussed on recently because various values of the density parameter involved seem to have different, rather concrete geometrical meanings. One variant of the density model is the *triangular model* (§ I.3.g.), which is somehow "less quotiented" and often produces only free groups.

All of these models lead to the same conclusion that a typical (finitely presented) group is hyperbolic. (For hyperbolic groups we refer the reader to the *Primer to geometric group theory* or to [BH99, Sho91a, GhH90, Ghy90, CDP90, Gro87].)

The topological approach of the *space of marked groups* (§ I.4.), though not a model of random groups itself, may be a nice framework to interpret some of these results in.

[2]It must be stressed that although any group has some presentation (as described in the *Primer to geometric group theory*), group presentations are mainly relevant for countable groups only, and the geometric methods work best for *finitely presented* groups. Thus the models of random groups currently used focus on those.

Another would-be model, arguably the most natural of all, the *temperature model*, is kind of a density model at all relator lengths simultaneously, thus producing non-finitely presented groups. It is addressed as an open question in § IV.k. since there still are no results about it.

Partially random groups. There has been some confusion due to the fact that the most famous (up to date) "random groups", those constructed by Gromov having no uniform embedding into the Hilbert space, exhibit quite different properties from what is hereafter described as typical for a random group (e.g. they are not hyperbolic). Actually the construction of these groups (thoroughly discussed in § III.) involves both random ingredients and manipulations quite specific to the goal of controlling uniform embeddings, and they are thus rather non-typical; dubbing them "partially random" would be more appropriate.

I.1. Forerunners: few-relator models

The statement that most groups are hyperbolic is *statistical*. It means that out of all possible group presentations, asymptotically most of them define hyperbolic groups. Here the asymptotics are taken with respect to the length of the relators involved.

Maybe the simplest statement expressing the overwhelming weight of hyperbolic presentations consists in considering the set of all presentations with a fixed number of relators and a bounded relator length, as in the following model.

Definition 1 (Few-relator model of random groups).
Let $\mathcal{R}_{k,\ell}$ be the set of all group presentations with k relators of length at most ℓ

$$\mathcal{R}_{k,\ell} = \{\langle\, a_1, \dots, a_m \mid r_1, \dots, r_k \,\rangle\,, \quad r_i \text{ reduced, } |r_i| \leqslant \ell \ \ \forall i\}$$

Let P be a property of a presentation. We say that P occurs with overwhelming probability *in this model if the share of presentations in $\mathcal{R}_{k,\ell}$ which have property P tends to 1 as ℓ tends to infinity.*

The following proposition was more or less implicit in the original formulation of small cancellation theory. Let us simply recall that $C'(\lambda)$ for $\lambda > 0$ is the condition that no two relators in a presentation share a common subword of length at least λ times the infimum of their lengths (we refer to [LS77] for small cancellation theory). When $\lambda < 1/6$ this implies hyperbolicity ([Gro87], 0.2.A).

Proposition 2.
For any k, for any $\lambda > 0$, the $C'(\lambda)$ small cancellation property occurs with overwhelming probability in the few-relator model of random groups. In particular, hyperbolicity occurs with overwhelming probability in this model, as well as torsion-freeness and cohomological dimension 2.

Of course, the overwhelming probability depends on k and λ: for very small λ's, it is necessary to take larger ℓ for the share to become close to 1.

Note also that since the number of possible relators of length ℓ grows exponentially with ℓ, a random relator of length at most ℓ actually has length between $\ell(1-\varepsilon)$ and ℓ, so that in this model all relators have almost the same length.

This proposition appears in [Gro87], 0.2.A (in the notation thereof, this is the case when ℓ_2/ℓ_1 is very close to 1). This is the model referred to as "généricité faible" (weak genericity) in [Ch91, Ch95]. The proof is straightforward. Take e.g. $k = 2$. The number of couples of reduced relators of length at most ℓ behaves like $(2m-1)^{2\ell}$, whereas the number of couples of relators sharing a common subword of length $\lambda\ell$ behaves roughly like $(2m-1)^{2\ell-\lambda\ell}$. So the share of couples of relators having a common subword of length $\lambda\ell$, for some $\lambda > 0$, is exponentially small when $\ell \to \infty$, so that the $C'(\lambda)$ condition is satisfied (a little more care is needed to treat the case of a piece between two parts of the same relator).

Remark 3.
The few-relator model of random groups appears as the 0-density case of the density model.

For this reason, results known to hold in this model are discussed below in § I.2.

In [Gro87], 0.2.A, Gromov immediately notes that it is not necessary to assume that all relators have lengths of the same order of magnitude to get hyperbolicity. This yields to the next model, which is technically much more difficult.

Definition 4 (Few-relator model with various lengths).
Given k integers $\ell_1, \ldots, \ell_k$, let

$$\mathcal{R}_{k,\ell_1,\ldots,\ell_k} = \{\langle\, a_1, \ldots, a_m \mid r_1, \ldots, r_k \,\rangle\,,\quad r_i \text{ reduced},\ |r_i| = \ell_i\}$$

be the set of presentations with k relators of prescribed lengths.

Let P be a property of a presentation. We say that P occurs with overwhelming probability *in this model if for any $\varepsilon > 0$ there exists an ℓ such that, if $\min \ell_i \geqslant \ell$, then the share of presentations in $\mathcal{R}_{k,\ell_1,\ldots,\ell_k}$ which have property P is greater than $1-\varepsilon$.*

In general, the small cancellation condition $C'(\lambda)$ is *not* satisfied in this model. Indeed, as soon as e.g. ℓ_2 is exponentially larger than ℓ_1, very probably the relator r_1 will appear as a subword of the relator r_2.

Once again however, hyperbolicity occurs with overwhelming probability. This is stated without proof in [Gro87], 0.2.A, and is referred to as "Théorème sans preuve" in [GhH90]. A little bit later, proofs were given independently by Champetier [Ch91, Ch95] (in the case $k = 2$) and Ol'shanskiĭ [Ols92].

Theorem 5.
With overwhelming probability, a random group in the few-relator model with various lengths is non-elementary hyperbolic, torsion-free, of cohomological dimension at most 2.

A few more results are known in this model. For $k = 2$, the boundary of the group is a Menger curve [Ch95] (see also § I.3.d.). Also the rank is the one expected [CS98] (and in particular the cohomological dimension in the previous theorem is actually 2):

Theorem 6.
With overwhelming probability, a random group in the few-relator model with various lengths has the following property: the subgroup generated by any $m-1$ generators chosen among $a_1, \ldots, a_m$ is free of rank $m-1$.

Moreover, thanks to a theorem of Champetier [Ch93] the spectral radius of the random walk operator associated with $a_1, \ldots, a_m$ (see definition in § I.3.f.) is arbitrarily close to the smallest possible value $\sqrt{2m-1}/m$ [CS98]. Using similar spectral bounds, it is proven in [CV96] that for $k=1$ (one relator), the semi-group generated by $a_1, \ldots, a_m$ is free.

Let us stress that contrary to the few-relator, one-length model, the few-relator model with various lengths is *not* subsumed in the density model below. It might, however, be recovered as an iterated random quotient at density 0 (see § II.2.), but the technical details needed to get this are still unclear.

I.2. Gromov's density

I.2.a. Definition of density. By the time Champetier and Ol'shanskiĭ had proven his first statement, Gromov had already invented another model, the *density model* ([Gro93], Chapter 9 entitled *Finitely presented groups: density of random groups and other speculations*). A continuous density parameter now controls the quantity of relators put in the random presentation. The sharpness of the notion is revealed through a phase transition theorem: if density is less than 1/2, then the random group is very probably infinite hyperbolic, whereas it is trivial at densities above 1/2.

Definition 7 (Density model of random groups).
Let F_m be the free group on $m \geqslant 2$ generators $a_1, \ldots, a_m$. For any integer ℓ let $S_\ell \subset F_m$ be the set of reduced words of length ℓ in these generators.

Let $0 \leqslant d \leqslant 1$. A random set of relators at density d, at length ℓ *is a $(2m-1)^{d\ell}$-tuple of elements of S_ℓ, randomly picked among all elements of S_ℓ (uniformly and independently).*

A random group at density d, at length ℓ *is the group G presented by $\langle\, a_1, \ldots, a_m \mid R \,\rangle$ where R is a random set of relators at density d, at length ℓ.*

We say that a property of R, or of G, occurs with overwhelming probability *at density d if its probability of occurrence tends to 1 as $\ell \to \infty$, for fixed d.*

Remark 8.
Slight variants of this historical definition exist, sometimes leading to more nicely expressed statements, e.g. replacing the sphere S_ℓ with the ball B_ℓ of words of length at most ℓ. They are discussed in § I.2.c.

Of course, the main point in this definition is the number $(2m-1)^{d\ell}$ of relators taken, which is actually quite large. Note that the set S_ℓ contains about $(2m-1)^\ell$ words, so that density is measured logarithmically (a fact we will meet again in § II.). The intuition behind this and the strong analogy with usual dimension and intersection theory are very nicely explained in [Gro93] (see also [Ghy03]): for a finite set X, the density of $A \subset X$ defined as $d(A) = \log \#A / \log \#X$ has, for "generic" A, lots of expected properties of a dimension (e.g. for the dimension of an intersection).

The basic idea is that $d\ell$ is the "dimension" of the random set of relators R (the set S_ℓ itself being considered of dimension ℓ because we have ℓ independent letter choices to make to specify an element in S_ℓ).

Classically, the dimension of a set (subspace in a vector space, algebraic submanifold) is the maximal number of "independent equations" that we can impose so that there still exists an element in the set satisfying them. For words, an "equation" will mean prescribing some letter in the word. Now consider e.g. a set of $2^{d\ell}$ random words of length ℓ in the two letters a and b; a simple counting argument shows that very probably, one of these words will begin with roughly $d\ell$ letters a (but not much more), meaning that this random set has "dimension" $d\ell$. More precisely:

Proposition 9.
Let R be a random set of relators at density d, at length ℓ. Let $0 \leqslant \alpha < d$. Then with overwhelming probability the following occurs: Any reduced word of length $\alpha\ell$ appears as a subword of some word in R.

Note that by a trivial cardinality argument, if $\alpha > d$ there exists a reduced word of length $\alpha\ell$ not appearing as a subword of any word in R.

Let us show on another example the strength (and correctness) of dimensional reasoning: Let us compute the probability that there exist two relators in R sharing a common subword of length $\alpha\ell$. The dimension of R is $d\ell$, so that the dimension of the set of couples $R \times R$ is $2d\ell$. Now sharing a common subword of length L imposes L equations, so that the "dimension" of the set of couples of relators sharing a common subword of length $\alpha\ell$ is $2d\ell - \alpha\ell$. So if $d < \alpha/2$ this dimension will tend to $-\infty$ as $\ell \to \infty$, implying that there will be no such couple of relators; conversely if $d > \alpha/2$ there will exist such a couple because dimension will be positive. What we have "shown" is:

Proposition 10.
Let $\alpha > 0$ and $d < \alpha/2$. Then with overwhelming probability, a random set of relators at density d satisfies the $C'(\alpha)$ small cancellation condition.

Conversely, if $d > \alpha/2$, then with overwhelming probability a random set of relators at density d does not satisfy the $C'(\alpha)$ small cancellation condition.

A rigorous proof ([Gro93], 9.B) is obtained by a simple counting argument, which in fact amounts to raising $(2m-1)$ to the exponents given by the various "dimensions" of the sets involved.

I.2.b. The phase transition. The striking phase transition theorem then proven by Gromov in [Gro93] is as follows.

Theorem 11.
Let G be a random group at density d.

- *If $d < 1/2$, then with overwhelming probability G is infinite, hyperbolic, torsion-free, of geometric dimension* 2.
- *If $d > 1/2$, then with overwhelming probability G is either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$.*

We include a proof of this theorem in § V. The argument appears in [Gro93], pp. 273–275; it suffers from omission of the case when a van Kampen diagram comprises the same relator several times. The proof of a similar-looking statement (Theorem 29 in the triangular model, see § I.3.g.) in [Żuk03] suffers from a similar but more subtle flaw (see § V.). A somewhat lengthy proof is given in [Oll04].

This calls for a few comments: Of course $\mathbb{Z}/2\mathbb{Z}$ occurs for even ℓ. What happens at exactly $d = 1/2$ is unknown and even the right way of asking the question is unclear (see § IV.a. for an elaboration on this). Note that Proposition 10 already implies the conclusion for $d < 1/12$, for then the presentation satisfies the good old $C'(1/6)$ small cancellation condition.

This theorem generalizes to random quotients of torsion-free hyperbolic groups (§ II.1. and § II.2.).

The reason for density 1/2 is the following: Recall the *probabilistic pigeon-hole principle*[3], which states that if in N holes we put much more than $\sqrt{N}$ pigeons then we will put two pigeons in the same hole (very probably as $N \to \infty$, provided that the assignment was made at random). In other words, a generic set of density more than 1/2 does self-intersect.

Density 1/2 is the case when the cardinal of the set of relators R is more than the square root of the cardinal of the set S_ℓ of words of length ℓ. In particular, this means that we very probably picked twice the same relator in R. *A fortiori*, very probably there are two relators $r_1, r_2 \in R$ differing just at one position i.e. $r_1 = wa_i^{\pm 1}$, $r_2 = wa_j^{\pm 1}$ with $|w| = \ell - 1$. But in the group $G = F_m/\langle R\rangle$, we have by definition $r_1 =_G r_2 =_G e$, so that $a_i^{\pm 1} =_G a_j^{\pm 1}$. Since there are only a finite number of generators, this will eventually occur for every value of i and j and every sign of the exponent, so that in G any generator will be equal to any other and to its inverse, implying that the group has only one or two elements.

This proves the trivial part of the theorem.

I.2.c. Variations on the model. Several points in the definition above are left for interpretation. First, let us stress that it is not crucial to take relators of length exactly ℓ: choosing lengths between ℓ and $\ell + o(\ell)$ would do as well. This even has several advantages: it kills the odd $\mathbb{Z}/2\mathbb{Z}$ in Theorem 11 and avoids matters of divisibility by 3 in the property (T) theorem (Theorem 27).

Actually the most natural setting is perhaps to choose at random words of length *at most* ℓ ("ball variant") instead of exactly ℓ. Since the number of words grows exponentially, most words so taken will be of length close to ℓ, but since the number

[3] A.k.a. the *birthday paradox*: in a class of more than 23 pupils there is a good chance that two of them share the same birthday. This is a simple combinatorial exercise.

of words taken is exponential too, some words will be shorter (at density d the shortest word will have length approximately $(1-d)\ell$). This variant simplifies the statements of Theorems 11, 27 and 38, is more natural for random quotients (§ II.), and the validity of all random group theorems proven so far seems to be preserved. In this text we chose to keep the historical Definition 7, in order to quote the literature without change; but e.g. for a textbook on random groups, the ball variant might be preferable.

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, so both interpretations are valid.

Numbers such as $(2m-1)^{d\ell}$ are not necessarily integers. We can either take the integer part, or choose two constants C_1 and C_2 and consider taking any number of words between $C_1(2m-1)^{d\ell}$ and $C_2(2m-1)^{d\ell}$. Once more this does not affect our statements at all.

One may hesitate between choosing reduced or cyclically reduced words. Once again this does not matter.

Section 4 of [Oll04] (in particular Remark 8) contains an axiomatic framework which allows to handle such a loose model and not to reprove all the theorems for each variant.

In all theorems stated in this text, not only does "with overwhelming probability" mean that the share of groups not having the property under consideration tends to 0 as $\ell \to \infty$, but actually the decay is exponential, that is, there exists a constant c (depending on everything except ℓ) such that this share is less than $\exp(-c\ell)$.

A very natural generalization of the density model is the temperature model, described in § IV.k.

Remark 12 (on density 0).
The intuition makes it clear that the only thing that matters is the exponent of growth of the number of relators. Thus, although it would follow from Definition 7 that a random set of relators at density 0 consists of one relator, we often use "density 0" to refer to a situation when the number of relators grows subexponentially with their lengths, e.g. the case of a constant number of relators (the few-relator model of Def. 1—but not the one of Def. 4).

I.3. Critical densities for various properties

A bunch of properties are now known to hold for random groups. This ranges from group combinatorics (small cancellation properties) to algebra (freeness of subgroups) to geometry (boundary at infinity, growth exponent, CAT(0)-ness) to probability (random walk in the group) to representation theory on the Hilbert space (property (T), Haagerup property). Some of the properties studied here are intrinsic to the group, others depend on a marked set of generators or on the standard presentation through which the random group was obtained.

Most interesting is the fact that some intrinsic properties vary with density (property (T), Haagerup property), thus proving that different densities can provide non-isomorphic groups (see § IV.b. for a discussion of this problem).

I.3.a. Van Kampen diagrams and small cancellation properties. These are the most immediate properties one gets for a random group. They are properties of the presentation, not of the abstract group.

Hyperbolicity of random groups is proven through isoperimetry of van Kampen diagrams (see the *Primer to geometric group theory* for what we need on van Kampen diagrams or [LS77, Ols91a, Rot95] for definitions and [Sho91a] for the link with hyperbolicity). Various, closely related formulations of this inequality for random groups appear in [Gro93, Ch91, Ols92, Ch95, Oll04, Oll-f]. We give the most recent one from [Oll-f], which is sharp and, combined with a result in [Ch94], gives a nice estimate for the hyperbolicity constant:

Theorem 13.
For every $\varepsilon > 0$, with overwhelming probability, every reduced van Kampen diagram D in a random group at density $d < 1/2$, at length ℓ satisfies

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

where $|\partial D|$ is the boundary length and $|D|$ the number of faces of D.

Consequently, the hyperbolicity constant δ of the random group satisfies

$$\delta \leqslant \frac{4\ell}{1 - 2d}$$

and the length of the smallest relation in the group is at least $\ell(1 - 2d - \varepsilon)$.

These properties are understood with respect to the standard presentation from which the random group was obtained.

This theorem of course implies Theorem 11. The complete argument is given in § V.

The isoperimetric constant is optimal in the sense that, with overwhelming probability, there exists a two-face van Kampen diagram D satisfying $|\partial D| \leqslant (1 - 2d + \varepsilon)\, \ell \,|D|$, which is just the failure of the $C'(2d+\varepsilon)$ small cancellation property (Prop. 10). For the hyperbolicity constant, clearly ℓ is the right order of magnitude but the real dependency on d when $d \to 1/2$ is unclear.

Next come some small cancellation conditions. By the way, actually as d approaches $1/2$, we have arbitrarily *large* cancellation (which refutes the expression "small cancellation on average" sometimes applied to this theory—we indeed measure cancellation on average, but it is not small), as results from the next proposition.

Recall that, given a group presentation, a *piece* is a word which appears as a subword of two different relators in the presentation, or as a subword at two different positions in the same relator (relators are considered as cyclic words and up to inversion). For $\alpha > 0$, the most often used $C'(\alpha)$ condition states that the length of any piece is less than α times the infimum of the lengths of the relators on which it appears. For an integer p, the $C(p)$ condition holds if no relator is the union of less than p pieces. The $B(2p)$ condition holds if the union of p consecutive pieces always makes less than half a relator. We have the implications $C'(1/2p) \Rightarrow B(2p) \Rightarrow C(2p{+}1)$. The $T(p)$ small cancellation condition is totally irrelevant for groups with lots of relators.

Conditions $C'(1/6)$, $C(7)$ or $B(6)$ imply hyperbolicity. Elementary counting arguments [OW-b] yield the following more erudite version of Proposition 10:

Proposition 14.
With overwhelming probability:

- *The $C'(\alpha)$ condition occurs if $d < \alpha/2$ and fails for $d > \alpha/2$,*
- *The $C(p)$ condition occurs if $d < 1/p$ and fails for $d > 1/p$,*
- *The $B(2p)$ condition occurs if $d < 1/(2p+2)$ and fails for $d > 1/(2p+2)$.*

These conditions being understood for the standard presentation from which the random group was obtained.

In particular, using the $C(7)$ condition, this proposition proves Theorem 11 up to density 1/7.

Closely related to small cancellation are Dehn's algorithm (see [LS77]), which holds for a group presentation when every reduced cyclic word representing the identity in the group has a subword which is more than half a subword of a relator in the presentation; and its stronger version, Greendlinger's Lemma (see [LS77] too), which holds when every reduced van Kampen diagram with at least two faces has at least two faces having more than half their length on the boundary of the diagram (in one piece). Every hyperbolic group admits some finite presentation satisfying Dehn's algorithm ([Sho91a], Theorem 2.12). However, for the standard presentation of a random group, a phase transition occurs at 1/5 [Oll-f]:

Theorem 15.
With overwhelming probability, if $d < 1/5$ the standard presentation of a random group at density d satisfies Dehn's algorithm and Greendlinger's Lemma. If $d > 1/5$, with overwhelming probability it does not satisfy any of the two.

This property is the last remnant of combinatorial small cancellation when density increases. It is crucial in the proof of Theorems 32 and 33 about action on a cube complex and failure of property (T).

I.3.b. Dimension of the group. A consequence of the isoperimetric inequality holding for any reduced van Kampen diagram is that the Cayley 2-complex associated with the presentation is aspherical [Gro93, Oll04], so that the group has geometric (hence cohomological) dimension 2 as stated in Theorem 11. The Euler characteristic of the group is thus simply $1 - m + (2m-1)^{d\ell}$.

In particular, since this Euler characteristic is positive for large ℓ, we get the following quite expected property (at least for $d > 0$, but this also holds at density 0 thanks to Theorem 18):

Proposition 16.
With overwhelming probability, a random group in the density model is not free.

Consideration of the Euler characteristic also implies that, for fixed m, the "dimension" $d\ell$ of the set of relations of the group is well-defined by its algebraic structure.

I.3.c. Algebraic properties at density 0: rank, free subgroups. When density is 0 (i.e. in the few-relator model, see Def. 1 and Remark 12), random groups keep lots of algebraic properties of a free group. In a certain sense, there are "no more" relations holding in the group than those explicitly put in the presentation. Several theorems in this direction are proven by Arzhantseva and Ol'shanskiĭ, using a technique of representation of subgroups of a group by labelled graphs (or finite automata) introduced by Stallings [Sta83], technique which will be discussed more in § III.1. Arzhantseva and Ol'shanskiĭ are able to extract, from failure of freeness in subgroups of a random group, a "too small" representation of one of the relators, which never occurs for random relators.

It is clear that some (all?) of these theorems do not hold at all densities. But they probably extend to small positive values of d, the determination of which is an interesting problem.

The first such theorem [AO96] is the following:

Theorem 17.
With overwhelming probability in the few-relator model of random groups with m generators and n relators, any subgroup generated by $m-1$ elements is free.

This is not true at all densities (see § IV.d.). When the $m-1$ generators of the subgroup are chosen among the m standard generators of the group, this is a particular case of Theorem 6.

The group itself is not free and more precisely [Ar97]:

Theorem 18.
In the few-relator model of random groups with $n \geqslant 1$ relators, no finite-index subgroup of the group is free.

As a corollary of these two theorems, we see that the rank of the random group in the few-relator model is exactly m, which, once again, does not hold at all densities (cf. § IV.d.).

Reusing the methods of Arzhantseva and Ol'shanskiĭ, Kapovich and Schupp prove that there is "only one" m-tuple generating the group. Recall [LS77] that for a m-tuple of elements $(g_1, \ldots, g_m)$ in a group, a Nielsen move consists in replacing some g_i with its inverse, or interchanging two g_i's, or replacing some g_i with $g_i g_j$ for some $i \neq j$. Obviously these moves do not change the subgroup generated by the m-tuple. The theorem [KS05] reads:

Theorem 19.
With overwhelming probability, in a random few-relator group G, any m-tuple of elements generating a non-free subgroup is Nielsen-equivalent in G to the standard m-tuple of generators w.r.t. which the random presentation was taken.

In particular, any automorphism of G lifts to an automorphism of F_m.

More properties of free groups are kept by random few-relator groups. In a free group, any subgroup is free; any finitely generated subgroup is quasiconvex; any non-trivial finitely generated normal subgroup has finite index ([LS77], Prop. I.3.12); the intersection of any two finitely generated subgroups is finitely generated (Howson's Theorem, [LS77], Prop. I.3.13). These properties are generalized as follows in [Ar97, Ar98]:

Theorem 20.
Let $L \geqslant 1$ be an integer. With overwhelming probability, a random few-relator group satisfies the following properties.

- *Any subgroup of rank at most L and of infinite index is free.*
- *Any subgroup of rank at most L is quasiconvex.*
- *Any non-trivial normal subgroup of rank at most L has finite index.*
- *The intersection of any two subgroups of rank at most L is quasiconvex and finitely generated.*

The overwhelming probability depends on L. For example, it is not clear whether all infinite-index subgroups are free or not.

The last point follows from the second one, noting that, in a finitely generated group, the intersection of two quasiconvex subgroups is quasiconvex and quasiconvex subgroups are finitely generated (see the nice [Sho91b]).

The results mentioned so far deal with properties of subgroups in the random group. One can wonder how subgroups of the free group are mapped to the random group. A theorem in this direction is the following [Ar00]:

Theorem 21.
Let $h_1, \ldots, h_k$ be elements of the free group F_m generating a subgroup H of infinite index. Then with overwhelming probability, the map from F_m to a random few-relator group is injective on H.

Conversely, it is easily seen that subgroups of finite index do not embed.

Of course this holds for elements $h_1, \ldots, h_k$ *fixed in advance*: it cannot be true that the quotient map is injective on all subgroups...

I.3.d. Boundary and geometric properties of the Cayley graph. We refer to [GhH90, CDP90, BH99] for the notion of boundary of a hyperbolic space.

Since the dimension of a random group is 2 (§ I.3.b.), by Corollary 1.4 of [BM91], the dimension of its boundary is 1. Champetier ([Ch95], Theorem 4.18) proves that at small density, the boundary is the most general object of dimension 1:

Theorem 22.
Let $d < 1/24$. Then with overwhelming probability, the boundary of a random group is a Menger curve. In particular the group is one-ended.

The bound 1/24 is probably not sharp. Let us recall that the Menger curve is the universal object in the category of compact metric spaces of dimension 1, see [And58]; it is (almost) characterized as the 1-dimensional, locally connected, locally non-planar continuum without local cut points (the boundary of a one-ended hyperbolic group never has cut points [Swa96]).

One-endedness probably holds at any density, but between 1/24 and 1/3 no simple criterion seems to apply. For $d > 1/3$, using Serre's theory of groups acting on trees [Ser77, HV89] it is a corollary of Theorem 27 on property (T):

Proposition 23.
Let $d > 1/3$. Then with overwhelming probability, a random group is one-ended.

At density 0, the Cayley graph of the group is not planar [AC04] (planarity of Cayley graph and complexes is an old story, see discussion in [AC04]). The result actually holds for generic $C'(1/8)$ small cancellation groups and so:

Theorem 24.
Let $d < 1/16$. With overwhelming probability, the Cayley graph (w.r.t. the standard generating set) of a random group at density d is not planar.

Actually the technique used in [AC04] allows to embed subdivisions of lots of finite graphs into the Cayley graph of a small-density random group.

I.3.e. Growth exponent. The growth exponent of a group presentation $G = \langle a_1, \ldots, a_m \mid R \rangle$ measures the rate of growth of balls in the group. Let B_L be the set of elements of the group G which can be written as a word of length at most L in the generators $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$. If G is the free group F_m, the number of elements of B_L is $1 + \sum_{k=1}^{L} (2m)(2m-1)^{L-1}$ which is the number of elements at distance at most L from the origin in the valency-$2m$ regular tree. The thing that matters here is the exponential growth rate of the balls,

$$g = \lim_{L \to \infty} \frac{1}{L} \log_{2m-1} \# B_L$$

(the limit exists thanks to the relation $\# B_{L+L'} \leqslant \# B_L \, \# B_{L'}$). This quantity is the *growth exponent* of the group G w.r.t. the generating set $a_1, \ldots, a_m$. It is at most 1, and equal to 1 if and only if G is the free group F_m on these generators. See [Har00] (chapters VI and VII), [GH97] or [Ver00] for some surveys and applications related to growth of groups.

Actually, the growth exponent of a random group at any density $d < 1/2$ is arbitrarily close to that of a free group with the same number of generators [Oll-b]. Of course, by Theorem 13 a random group behaves like a free group up to scales $\ell(1-2d)$, but growth is an asymptotic invariant taking into account the non-trivial geometry of the group at scale ℓ, so it is somewhat surprising that the growth exponent is large independently of the density d (except if $d > 1/2$ where of course it drops to 0). Computing the growth exponent was initially an attempt to build a continuous quantity depending on density.

Theorem 25.
Let $\varepsilon > 0$ and $0 \leqslant d < 1/2$. Then with overwhelming probability, the growth exponent of a random group at density d lies in the interval $[1-\varepsilon; 1)$ (w.r.t. the standard generating set).

Note that non-sharp bounds for the growth exponent can be obtained from the spectral estimates discussed in § I.3.f. (see discussion in [Oll-b]).

When $d < 1/12$ the random group satisfies the $C'(1/6)$ small cancellation condition, and in this case this result is related to a theorem of Shukhov [Shu99] stating that $C'(1/6)$ groups with long enough relators and "not too many" relators have growth exponent close to 1. Shukhov's "not too many" relators condition is strikingly reminiscent of a density condition.

I.3.f. Random walk in a typical group. A group together with a generating set defines a random walk, which consists in starting at e and, at each step, multiplying by one of the generators or its inverse, chosen at random (this is the simple random walk on the Cayley graph). A foundational paper of this theory is that of Kesten [Kes59], see also [Gri80, GH97, Woe00] for some reviews.

One quantity containing a lot of information about the random walk is the spectral radius of the random walk operator (see [Kes59]). Let P_t be the probability that at time t, the random walk starting at e is back at e. The *spectral radius* of the random walk is defined as

$$\rho = \lim_{\substack{t\to\infty \\ t \text{ even}}} (P_t)^{1/t}$$

(the limit exists thanks to the property $P_{t+t'} \geqslant P_t P_{t'}$). One restricts oneself to even t because there might be no odd-length path from e to e in the Cayley graph. This quantity is at most 1 (a value achieved if and only if the group is amenable [Kes59]), and at least $\frac{\sqrt{2m-1}}{m}$ where m is the number of generators (achieved if and only if the group is free on these generators).

Just as for the growth exponent, it came out as a surprising fact that the spectral radius of the random walk on a random group does not depend on density [Oll05a], except of course when $d > 1/2$ where it suddenly jumps to 1. Once again this cannot be interpreted simply by saying that random groups are free up to scale $\ell(1-2d)$, because the spectral radius is an asymptotic invariant taking into account the non-trivial geometry at scale ℓ.

Theorem 26.
Let $\varepsilon > 0$ and $0 \leqslant d < 1/2$. Let $\rho(F_m) = \frac{\sqrt{2m-1}}{m}$ be the spectral radius of the random walk on the free group F_m.

Then with overwhelming probability, the spectral radius of the random walk on a random group at density d lies in the interval $(\rho(F_m); \rho(F_m)+\varepsilon)$.

At density 0 this follows from a theorem of Champetier [Ch93], which, as mentioned earlier (§ I.1.), also holds in the few-relator model with various lengths.

Consequently, the growth exponent of the kernel of the map $F_m \to G$ (the *cogrowth exponent* of G) is less than $1/2+\varepsilon$, thanks to a formula by Grigorchuk ([Gri80, Ch93, Oll04]). This answers Gromov's question 9.B.(c) in [Gro93]: normal closures of random sets of density $< 1/2$ generated by random large elements in F_m have "density" (growth exponent) less than $1/2+\varepsilon$.

This result also has a nice interpretation in terms of a random walk on an infinite tree with lots of "zero-length" bridges added at random (in an "equivariant" way). Indeed the random walk on a random group G can be thought of as a random walk on the Cayley graph of the free group F_m where elements of F_m mapping to the same element of G are linked by a bridge through which the random walk can "instantly" travel. The theorem then states that adding "lots of" bridges equivariantly does not change the probability to come back to (a point connected by bridges to) the origin, up till some density when suddenly any point is connected to the origin by a sequence of bridges.

I.3.g. Property (T) and the triangular model. Kazhdan's property (T) of a group has to do with the behavior of unitary actions of the group on the Hilbert space and basically asks that, if there are unitary vectors which the group action moves by arbitrarily small amounts, then there is a vector fixed by the action. It has proven to be linked with numerous algebraic or geometric properties of the group. We refer to [HV89, BHV, Val02a] for reviews and basic properties.

The so-called *spectral criterion* is a sufficient condition for property (T) of a discrete group, which is an explicitly checkable property of the ball of radius 1 in the Cayley graph w.r.t. some generating set. The neatest statement is to be found in [Żuk03], see also [Żuk96, BŚ97, Pan98, Wan98, Val02a]. Gromov (part 3 of [Gro03]) put this result in a more general context, which allowed Ghys to write a very simple proof [Ghy03, Oll-d].

It happens that in the density model, after suitable manipulations of the presentation, this criterion is satisfied as soon as $d > 1/3$. It is not known whether this latter value is optimal (compare Theorem 32 below and § IV.c.).

Theorem 27.
Let $d > 1/3$ and let G be a random group at density d and at lengths $\ell, \ell+1$ and $\ell+2$. Then, with overwhelming probability, G has property (T).

The necessity to take a random quotient at three lengths simultaneously is a technical annoyance due to the too restrictive definition of the density model, which disappears if we replace the sphere by the ball in Definition 7 (see Remark 8 and § I.2.c.). This results from the necessity to have some relators of length a multiple of 3, as we explain now.

This theorem is proven using an intermediate random group model better suited to apply the spectral criterion, the *triangular model*, which we now define. This model consists in taking relators of length only 3, but letting the number of distinct generators tend to infinity. Żuk [Żuk03] wrote a proof that property (T) holds in this model at density $d > 1/3$ (Theorem 31); it is then possible to carry the result to the density model (but actually Theorem 27 seems not to be written anywhere explicitly in the literature).

Definition 28 (Triangular model).
Let n be an integer and let $b_1, \ldots, b_n$ be n distinct symbols. Let $W_{n,3}$ be the set of words of length 3 in $b_1, \ldots, b_n, b_1^{\pm 1}, \ldots, b_n^{\pm 1}$ which are cyclically reduced.

Let $0 \leqslant d \leqslant 1$. A random group on n relators at density d in the triangular model *is the group presented by $G = \langle\, b_1, \ldots, b_n \mid R_3 \,\rangle$ where R_3 is a set of $(\#W_{n,3})^d$ words taken at random in $W_{n,3}$.*

A property of G is said to occur with overwhelming probability *in this model if its probability of occurrence tends to 1 as $n \to \infty$.*

Note that $\#W_{n,3} \sim (2n)^3$ for large n. The density intuition is the same as above: the number of relators taken is a power of the total number of possibilities.

Actually there is a natural homomorphism from a random group in the triangular model to a random group in the density model, at the same density. This goes as follows: Let m be the number of generators used in the density model, let ℓ be a length which is a multiple of 3, and let $W'_{m,\ell/3}$ be the set of all reduced words of length $\ell/3$ in the symbols $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$. Now take $n = \frac{1}{2}\#W'_{m,\ell/3}$ and define a map φ from the free group $\langle b_1, \ldots, b_n \rangle$ to the free group $\langle a_1, \ldots, a_m \rangle$ by enumerating all the words in $W'_{m,\ell/3}$ and sending each b_i to a distinct such word (and sending inverses to inverse words). Note that if $w \in W_{n,3}$ then $\varphi(w)$ is a word of length ℓ in the $a_i^{\pm 1}$'s.

Now if $G_3 = \langle\, b_1, \ldots, b_n \mid R_3 \,\rangle$ is a random group in the triangular model, we can define a group $G = \langle\, a_1, \ldots, a_m \mid \varphi(R_3) \,\rangle$, in which the relators will have length ℓ. If G_3 is taken at density d, then by definition it consists of $(\#W_{n,3})^d \sim (2n)^{3d}$ relators, so that $\#\varphi(R_3) = \#R_3 \sim (2n)^{3d} = (\#W'_{m,\ell/3})^{3d} \sim \left((2m-1)^{\ell/3}\right)^{3d} = (2m-1)^{3d\ell}$, in accordance with the density model. Note also that the image of the uniform law on $W_{n,3}$ is (almost) the uniform law on the set of reduced words of length ℓ in the $a_i^{\pm 1}$'s. (The "almost" comes from the fact that if $w = b_{i_1} b_{i_2} b_{i_3}$, then $\varphi(w)$ may not be reduced at the junction points of $\varphi(b_{i_j})$ with $\varphi(b_{i_{j+1}})$, but the density model is robust to such a slight modification, see § I.2.c.).

So, up to this latter technicality, there is a natural homomorphism $\varphi : G_3 \to G$ from a random group in the triangular model to a random group in the density model, *at the same density*. This means that the triangular model is "less quotiented" than the density one.

It is possible to prove [Żuk03] quite the same hyperbolicity theorem as for the density model:

Theorem 29.
If $d < 1/2$, then with overwhelming probability a random group in the triangular model, at density d, is non-elementary hyperbolic. If $d > 1/2$, it is trivial with overwhelming probability.

But the fact that groups in the triangular model are "larger" than those in the density model is especially clear when considering the following proposition.

Proposition 30.
If $d < 1/3$, then with overwhelming probability, a random group in the triangular model at density d is free.

Of course its rank is smaller than n. This results from the fact that at density $d < 1/3$ in the triangular model, the dual graph of any van Kampen diagram is a tree.

Żuk [Żuk03] wrote a proof that in the triangular model, the spectral criterion for property (T) is satisfied:

Theorem 31.
If $d > 1/3$, then with overwhelming probability, a random group in the triangular model at density d has property (T).

In the triangular model, density $1/3$ corresponds to the number of relators being equal to the number of generators. So typically at $d > 1/3$ every generator appears a large number of times in the relators, which is not the case for $d < 1/3$. Consequently, the link of e in the Cayley graph will be a random graph with a large number of edges per vertex. Such graphs have a very large (close to 1) spectral gap and the spectral criterion for property (T) mentioned above applies.

Actually, as an intermediate step Żuk uses yet another variant of the triangular model (based on random permutations, see section 7.1 in [Żuk03]), which is rather artificial for random groups but arises very naturally in the context of random graphs, a crucial tool of the proof. The transfer to the standard triangular model involves use of the matching theorem.

Now property (T) is stable under quotients. Using the morphism $\varphi : G_3 \to G$ above, Theorem 27 follows from Theorem 31 and from all the details we've omitted (e.g. the necessary modifications to make in order to get a surjective φ, or handling of the reduction problems). It seems that actually neither these details nor Theorem 27 itself do appear in the literature.

I.3.h. Testing the triangular model: Gromov vs. the computer. There is an amusing story to be told about the triangular model. In 2001, Richard Kenyon performed computer experiments to test Gromov's statement (Theorem 29). He used Derek Holt's KBMAG package [Hol95] to test triviality of random groups in the triangular model. The tests were made up to $n = 500$ generators using about 2000 relations (which makes d slightly above $1/3$). The results suggested that triviality occurred as soon as $d > 1/3$, in contradiction with Theorem 29. Kenyon subsequently reviewed Gromov's proof of Theorem 11 given in [Gro93] and pointed out the omission of van Kampen diagrams featuring the same relator several times (see § V.).

The author performed another series of experiments and analyzed the results by hand. It turned out that each time the group was trivial at $1/3 < d < 1/2$, this was due to some "exceptional event" whose asymptotic probability should be very small; but the combinatorial factor counting these events, although bounded, is quite large (a fact that may be related to the huge constants appearing in the local-global principle, see discussion in § V.); so the phenomenon should disappear when using larger n. On the other hand, it was not difficult to correct Gromov's argument (this led to the proof of Theorem 11 given in [Oll04]). The observed change of behavior of the algorithm at $d > 1/3$ might be related to Proposition 30: at $d < 1/3$ the group is non-trivial (actually free) for trivial reasons, whereas at $1/3 < d < 1/2$ the reasons for non-triviality involve the full strength of hyperbolic theory.

The triangular model has seemingly been less successful than the density model. Comparing Proposition 30 and Theorem 31, this model may be quite specific to the study of property (T) (but a one-step proof of Theorem 27 using the spectral criterion applied to a generating set made of words of length $\ell/3$ is feasible). Moreover the triangular model does not generalize to a theory of random quotients of given groups (§ II.1., § II.2.). On the contrary, in the usual model density controls the occurrence of several combinatorial and geometric events; we now turn to the description of some transformations happening at densities $1/6$ and $1/5$.

I.3.i. Cubical CAT(0)-ness and the Haagerup property. In view of Theorem 27, one can wonder whether $1/3$ is the optimal density value for the occurrence of property (T). It is not true that all random groups have this property: indeed, random groups at $d < 1/12$ are $C'(1/6)$ small cancellation groups, and, following Wise [Wis04], those do not have property (T). This happens to be the case up to density at least $1/5$ [OW-b]:

Theorem 32.
Let $d < 1/5$. Then with overwhelming probability, a random group at density d has a codimension-1 subgroup. In particular, it does not have property (T).

The codimension-1 subgroup (the existence of which excludes property (T) by a result in [NR98]) is constructed through a technique developed by Sageev [Sag95], extended among others by Wise [Wis04], related to actions of the group on cube complexes. When $d < 1/6$, the construction of [Wis04] fully applies and provides a complete geometrization theorem [OW-b]:

Theorem 33.
Let $d < 1/6$. Then with overwhelming probability, a random group at density d acts freely and cocompactly on a CAT(0) cube complex. Moreover it is a-T-menable (Haagerup property).

Like property (T), with which it is incompatible, the Haagerup property of a group has to do with its actions on the Hilbert space. It amounts to the existence of a proper isometric affine action on the Hilbert space, which is a kind of flexibility excluding property (T). We refer to [CCJJV01] for a fact sheet on the Haagerup property. For discrete groups, a very nice equivalent definition is the existence of a proper action on a space with measured walls [CMV04].

The strategy is to construct walls [HP98] in the group. Natural candidates to be walls are *hypergraphs* [Wis04], which are graphs built from the Cayley complex as follows: the vertices of the hypergraphs are midpoints of edges of the Cayley complex, and the edges of the hypergraphs connect vertices corresponding to midpoints of diametrally opposite edges in faces of the Cayley complex (assuming that all relators have even length).

Densities $1/5$ and $1/6$ come into play as follows. If the group presentation satisfies the conclusion of Theorem 15 (kind of Greendlinger's Lemma), which happens up to $1/5$, then the hypergraphs are trees embedded in the Cayley complex and so are genuine walls. The stabilizers of these walls provide codimension-1 subgroups, thus refuting property (T) [NR98]. On the contrary it happens that for $d > 1/5$, there is only one hypergraph, which passes through every edge of the Cayley complex [OW-b]...

Following Sageev's [Sag95] original ideas, there is a now standard [NR98, Nic04, Wis04, CN, HW] correspondence between spaces with walls and cubical complexes. In our case, below density $1/6$ (but not above) the hypergraphs have some convexity properties and moreover two given hypergraphs cannot intersect at more than one point (except for degenerate cases); this allows to show that there are "enough" walls for the cube complex construction to work [HW], getting a free, cocompact action of the random group. (It seems likely however that Theorem 33 holds up to density $1/5$.)

The Haagerup property follows either from consideration of the cube complex as in [NR98], or from general properties of groups acting on spaces with walls ([CMV04], after a remark of Haglund, Paulin and Valette).

According to Proposition 10, random groups at $d < 1/6$ satisfy the $C(6)$ small cancellation condition. So an interpretation of Theorem 33 is that "generic" $C(6)$ groups have the Haagerup property. It is currently an open question to know whether some $C(6)$ groups can have property (T).

This closes our journey through the influence of density on properties of the group. Space was missing to draw the geometric pictures corresponding to the events considered; but each time, density allows or forbids the existence of very concrete diagrams in the Cayley complex with certain metric properties relevant to the question under study.

I.4. The space of marked groups

A *group marked with m elements* is a finitely generated group G, together with an m-tuple of elements $g_1, \ldots, g_m \in G$ such that these elements generate G; or, equivalently, a group G together with a surjective homomorphism $F_m \to G$.

For fixed m, the space $\mathcal{G}_m$ of all groups marked with m elements has a natural topology, apparently first introduced in Grigorchuk's celebrated paper [Gri84], part 6 (also compare the end of [Gro81]): two marked groups $(G, (g_i))$ and $(G', (g'_i))$ are close if the kernels of the two maps $F_m \to G$, $F_m \to G'$ coincide in a large ball of F_m, or, equivalently, if some large balls in the Cayley graphs of G and G' w.r.t. the two given generating sets are identified by the mapping $g_i \mapsto g'_i$. This means that the two generating m-tuples have the same algebraic relations up to some large size.

We refer to [Pau03, Ch00, Gri84] for basic properties of the space $\mathcal{G}_m$. It is compact, totally discontinuous. Every finite group is an isolated point. The subspaces of Abelian groups, of torsion-free groups are closed. Finitely presented groups are dense in $\mathcal{G}_m$. Any finitely presented group has a neighborhood consisting only of quotients of it. The Minkowski dimension of $\mathcal{G}_m$ is infinite [Guy].

Isomorphism of groups defines a natural equivalence relation on $\mathcal{G}_m$. It happens that this relation is extremely irregular from a measurable point of view, so that it is not possible to measurably classify finitely generated groups by a real number [Ch91, Ch00]:

Theorem 34.
Let $m \geqslant 2$. There exists no Borel map $\mathcal{G}_m \to \mathbb{R}$ constant on the isomorphism classes and separating these classes.

Actually this equivalence relation is as irregular as a countable equivalence relation can be [TV99]. Let X and X' be Borel sets and let R, R' be Borel equivalence relations on X and X' respectively, with countable classes. Say that R is *reducible* to R' if there exists a Borel map $f : X \to X'$ such that $x\,R\,y \Leftrightarrow f(x)\,R'\,f(y)$. In other words, R' is more complex than R. The theorem reads [TV99]:

Theorem 35.
Any Borel equivalence relation with countable classes is reducible to the isomorphism relation on $\mathcal{G}_5$.

A review of the results and uses of this space is beyond the scope of our work. In fact, it happens that the closure of the isomorphism class of the free group (the *limit groups* of Sela [Sel]) is already quite complex [Sel, CG05]. We focus here on the aspects linked to the idea of typicality for groups.

The usual notion of topological genericity (G_δ-dense sets *à la* Baire) is not very interesting due to the totally discontinuous nature of the space; e.g. the set of Abelian groups is open-closed, as is any finite marked group alone, so that any Baire-generic property has to hold for these classes of groups. For these reasons, so far this space has not be used to define an alternate notion of a "typical" group competing with Gromov's probabilistic approach.

Nevertheless, following Champetier we can use our prior knowledge of genericity of torsion-free hyperbolic groups (Theorems 5 and 11) to restrict ourselves to the closure in $\mathcal{G}_m$ of those groups, and try to identify generic properties therein. Indeed this program happens to work very well [Ch91, Ch00]:

Theorem 36.
Let $\mathcal{H}_m^{\mathrm{tf}}$ be the closure in $\mathcal{G}_m$ of the subspace of all non-elementary, torsion-free hyperbolic groups. Then there is a G_δ-dense subset $X \subset \mathcal{H}_m^{tf}$ such that any group $G \in X$ satisfies the following properties:

- *Its isomorphism class is dense in $\mathcal{H}_m^{tf}$,*
- *It is torsion-free,*
- *It is of rank 2,*

- *It is of exponential growth, non-amenable,*
- *It contains no free subgroup of rank* 2,
- *It satisfies Kazhdan's property* (T),
- *It is perfect,*
- *It has no finite quotient but the trivial group.*

So all these properties can be viewed as generic properties of an "infinitely presented typical group".

Note however that such properties do depend on the class of groups we take the closure of. If, of Theorems 5 and 11, we had only retained the fact that a random group is non-elementary hyperbolic (and forget it is torsion-free), then we would naturally have considered the closure $\mathcal{H}_m$ in $\mathcal{G}_m$ of the subspace of non-elementary hyperbolic groups, in which case we get the following [Ch91, Ch00] (compare [Ols91c]):

Theorem 37.
There is a G_δ*-dense subset of* $\mathcal{H}_m$ *consisting only of groups which are infinite and all elements of which are of torsion.*

Infinite torsion groups have long been sought for (*Burnside problem* dating back to 1902). They were constructed for the first time in 1964 by Adyan and Novikov and have been the source of an abundant literature since then (see e.g. [Iva98, Gup89] for reviews). Diagrammatic methods for this problem were introduced by Ol'shanskiĭ [Ols82, Ols83]. It seems that hyperbolic groups are a natural way towards infinite torsion groups ([DG], [IO96], [Iva94], [Ols93], [Ols91c], chapter 12 of [GhH90], section 4.5.C of [Gro87]).

The strength of this topological approach compared to the probabilistic one is that it gives access to infinitely presented groups. The drawback is that it does not provide by itself a non-trivial notion of generic properties of groups: one has to combine it with prior knowledge from the probabilistic approach. Once properties known to hold with overwhelming probability for finitely presented groups are selected (and the result may depend on this choice), the closure of these groups in the space $\mathcal{G}_m$ provides a notion of genericity for infinitely presented groups. Of course the notions of genericity for finitely and infinitely presented groups cannot but be mutually incompatible.

Note however that the probabilistic approach does not provide a well-defined notion of "random group" either since one has to consider a family of probability measures indexed by the length ℓ of the relators; but at least this defines a notion of a generic property of finitely presented groups when $\ell \to \infty$.

The temperature model (§ IV.k.), if understood, would solve all these problems, noticeably by providing a natural family of (quasi-invariant?) probability measures on $\mathcal{G}_m$. See also § IV.g. for questions arising in this framework.

Part II.
Typical elements in a group

Quoting from [Gro87], 5.5F: "Everything boils down to showing that adding 'sufficiently random' relations to a non-elementary word hyperbolic group gives us a word hyperbolic group again[...]" Considering random elements in a given group is often a good way to embody the intuition of which properties are true when "nothing particular happens" and when the elements are "unrelated". The behavior of random elements is often the best possible.

We illustrate this on two categories of examples: The first is that typical elements in a (torsion-free) hyperbolic group can be killed without harming the group too much (robustness of hyperbolicity), and the probabilistic approach allows to quantify very precisely the number of elements that can be killed. The second is a sharp counting of the number of one-relator groups *up to isomorphism*, the idea being that a random relator is nicely behaved, implying a rigidity property, and that by definition typical relators are the most numerous so that it is enough to count only them.

For this review, typical elements in a group were considered only insofar as they are put in some group presentation and provide new, randomly-defined groups, hence the two topics selected. Random elements in a finite or infinite group have plenty of interesting properties by themselves, which are not covered here. See for example the nice [Dix02] for the case of finite groups.

II.1. Killing random elements of a group

II.1.a. Random quotients by elements in a ball. Theorem 11 states that a random quotient of the free group is hyperbolic. One can wonder whether a random quotient of an already hyperbolic group stays hyperbolic, and this is the case. In other words, hyperbolicity is not only generic but also robust. This is all the more reasonable as, from a geometric point of view, the intuition is that (torsion-free) hyperbolic groups supposedly behave like free groups.

The following is Theorem 3 of [Oll04] (up to the benign replacement of spheres by balls, see § I.2.c.), which generalizes the phase transition of Theorem 11 above. As usual "with overwhelming probability" means "with probability tending to 1 as $\ell \to \infty$".

Theorem 38.
Let G_0 be a non-elementary, torsion-free hyperbolic group equipped with some finite generating set, and let B_ℓ be the set of elements of G_0 with norm at most ℓ w.r.t. this generating set.

Let $0 \leqslant d \leqslant 1$. Let $R \subset G_0$ be a set obtained by picking at random $(\#B_\ell)^d$ times an element in B_ℓ. Let $G = G_0/\langle R\rangle$ be the random quotient obtained.

- *If $d < 1/2$, then with overwhelming probability G is (non-elementary) hyperbolic.*
- *If $d > 1/2$, then with overwhelming probability $G = \{e\}$.*

The explanation of the $1/2$ is of course exactly the same as for Theorem 11, namely the probabilistic pigeon-hole principle, although the proof for $d < 1/2$ is considerably more difficult. Note again that the number $(\#B_\ell)^d$ is rather large. This phenomenon seems to be quite robust and general and might be generalized to other subsets in which the generators are picked, and maybe other classes of groups (§ IV.f.).

The torsion-freeness assumption can be relaxed to a "harmless torsion" one, but it cannot be completely removed, the obstruction being growth of the centralizers of torsion elements (see Theorem 41 below, § IV.f. and [Oll05b]).

We refer to § IV.f. for natural questions and open problems concerning these random quotients.

From a constructive point of view, it might seem quite difficult to pick random elements uniformly distributed in the ball of a group, compared to the easy generation of random words as used in Theorems 11 and 40. However, algorithmic properties of hyperbolic groups are very nice: equality of two elements is decidable in real time [Hol00], every element can be efficiently [EH] put into a normal form, and there is an explicit finite automaton enumerating these normal forms ("Markov codings": section 5.2 of [Gro87], [GhH90]).

According to Gromov's quote above, the idea that unrelated elements in a hyperbolic group can be killed is quite old. In a deterministic context, this "relative small cancellation", presented in section 5.5 of [Gro87] (where Gromov refers to Ol'shanskiĭ's paper [Ols83]) was later formalized by Ol'shanskiĭ (section 4 of [Ols93]), Champetier [Ch94] and Delzant [Del96a]. This theory generalizes the usual small cancellation $C'(\lambda)$ to elements chosen in a hyperbolic group. But, just as usual small cancellation stops at density $1/12$ for random groups, relative small cancellation is too restrictive and does not make it up to the maximal number of elements one can kill, hence the interest of the random point of view.

II.1.b. Growth of random quotients. A theorem stating that the growth exponent does not change much under such a quotient, generalizing Theorem 25, has been proven [Oll-b] (we refer to § I.3.e. for the definition of the growth exponent). Note that by the results in [AL02], this exponent cannot stay unchanged.

Theorem 39.
Let G_0 be a non-elementary, torsion-free hyperbolic group generated by the finite set S. Let B_ℓ be the ball in G_0 and g the growth exponent of G_0, both w.r.t. S. Let G be a random quotient of G_0 by elements of B_ℓ at density d as in Theorem 38, and suppose of course $d < 1/2$.

Then, for any $\varepsilon > 0$, with overwhelming probability the growth exponent of G lies in the interval $(g - \varepsilon; g)$.

It is likely [Oll-e] that the spectral radius of the random walk operator on the group does not change much too (compare Theorem 42).

II.2. Killing random words, and iterated quotients

II.2.a. Random quotients by words. Theorem 38 describes what happens when quotienting a hyperbolic group by random elements in it. Another possible generalization of Theorem 11 is to quotient by random words in the generators. Though maybe not as intrinsic, this model has the advantage that the notion of random quotient becomes independent of the initial group (within the class of marked groups); in particular, it allows to study successive random quotients of a group taken w.r.t. one and the same generating set, as used notably in [Gro03].

Of course, the unavoidable consequence of the model being independent on the initial group is that the critical density will depend on this group. Actually the critical density is equal to the exponent of return to e of the simple random walk w.r.t. the generating set $a_1, \ldots, a_m$ considered: basically, if this probability behaves like $(2m)^{-\alpha t}$ for large times t, the critical density will be α. The result reads ([Oll04], Theorem 4):

Theorem 40.
Let G_0 be a torsion-free hyperbolic group generated by the elements $a_1, \ldots, a_m$.

Let $(w_t)_{t\in\mathbb{N}}$ be the trajectory of a simple random walk in G_0 w.r.t. the generators $a_i^{\pm 1}$ and let

$$d_{\text{crit}} = -\lim_{\substack{t\to\infty \\ t \text{ even}}} \frac{1}{t} \log_{2m} \Pr(w_t =_{G_0} e) = -\log_{2m} \rho(G_0)$$

where ρ is the spectral radius of the random walk operator ([Kes59] or § I.3.f.). Note that $d_{\text{crit}} > 0$ unless G_0 is elementary.

Let $0 \leqslant d \leqslant 1$ and let W_ℓ be the set of all $(2m)^\ell$ words of length ℓ in the $a_i^{\pm 1}$'s. Let R be the random set obtained by picking $(2m)^{d\ell}$ times at random a word in W_ℓ.

Let $G = G_0/\langle R\rangle$ be the random quotient so obtained. Then with overwhelming probability:

- *If $d < d_{\text{crit}}$, then G is (non-elementary) hyperbolic.*
- *If $d > d_{\text{crit}}$, then $G = \{e\}$.*

Once again the spirit of the density model is to kill a number of words equal to some power d of the total number of words $(2m)^\ell$. Note that $d_{\text{crit}} < 1/2$, even when G_0 is the free group (the difference with Theorem 11 being that we use plain words instead of reduced ones).

There is also a version of Theorem 40 using reduced words instead of plain words (thus a formal generalization of Theorem 11). In this version (Theorem 2 of [Oll04]), $2m$ is to be replaced with $2m-1$ everywhere and a $\mathbb{Z}/2\mathbb{Z}$ might appear for $d > d_{\text{crit}}$ and even ℓ. The critical density is now equal to $1/2$ for the free group, and to the exponent of return to e of the reduced random walk in G_0 (i.e. the random walk with no immediate backtracks) for a non-free hyperbolic group G_0, which is equal to 1 minus the cogrowth exponent of G_0 [Oll04].

Theorems 38 and 40 are of course not proven independently. Section 4 of [Oll04] extracts axioms under which quotients of a hyperbolic group by elements taken under some probability measure yield a hyperbolic group again. These axioms have to do with exponents of large deviations of the measure.

Lots of open problems concerning random quotients of hyperbolic groups are stated in § IV.f.

II.2.b. Harmful torsion. As briefly mentioned above, the torsion-freeness assumption can be relaxed to a "harmless torsion" one demanding that the centralizers of torsion elements are either finite, or virtually $\mathbb{Z}$, or the whole group [Oll04]. But in [Oll05b] we give an example of a hyperbolic group with "harmful" torsion, for which Theorem 40 does not hold; moreover its random quotients actually exhibit three genuinely different phases instead of the usual two.

Theorem 41.
Let $G_0 = (F_4 \times \mathbb{Z}/2\mathbb{Z}) \star F_4$ equipped with its natural generating set, where $\star$ denotes a free product. Let d_{crit} be defined as above. Then there exists a density $0 < d'_{\text{crit}} < d_{\text{crit}}$ such that quotients of G_0 by random words at density $d > d'_{\text{crit}}$ are trivial with overwhelming probability.

What happens is that above some density corresponding to the probability with which the random walk in G_0 sees the factor $F_4 \times \mathbb{Z}/2\mathbb{Z}$, the factor $\mathbb{Z}/2\mathbb{Z}$ becomes central in the random quotient, so that above this density random quotients of $(F_4 \times \mathbb{Z}/2\mathbb{Z}) \star F_4$ actually behave like random quotients of $F_8 \times \mathbb{Z}/2\mathbb{Z}$, which has a lower critical density.

A more careful analysis reveals the presence of two genuinely different phases for random quotients of $(F_4 \times \mathbb{Z}/2\mathbb{Z}) \star F_4$ in addition to the trivial phase, depending on whether or not the $\mathbb{Z}/2\mathbb{Z}$ factor is central in the quotient. Elaborating on this construction, hyperbolic groups with torsion whose random quotients exhibit more than three different phases can probably be built. We refer to [Oll05b] for the details.

II.2.c. Cogrowth of random quotients, and iterated quotients. Since the critical density of the initial group G_0 is controlled by the spectral radius of the random walk operator, one might wonder what is the new value of this spectral radius for the random quotient (in particular, if it stays small enough, then we can take a new random quotient at a larger length). The answer from [Oll05a] (using results of [Ch93]), generalizing Theorem 26, is that it stays almost unchanged:

Theorem 42.
Let G_0 be a torsion-free hyperbolic group generated by the elements $a_1, \ldots, a_m$. Let $\rho(G_0)$ be the spectral radius of the random walk operator on G_0 w.r.t. this generating set; let $d_{\text{crit}} = -\log_{2m} \rho(G_0)$ and let G be a quotient of G_0 by random words at density $d < d_{\text{crit}}$ as in Theorem 40.

Then, for any $\varepsilon > 0$, with overwhelming probability the spectral radius $\rho(G)$ of the random walk operator on G w.r.t. $a_1, \ldots, a_m$ lies in the interval $(\rho(G_0); \rho(G_0) + \varepsilon)$.

The same theorem holds for quotients by random reduced words, and, very likely [Oll-e], for quotients by random elements of the ball as in Theorem 38.

As a corollary, we get that the critical density for the new group G is arbitrarily close to that for G_0. So we could take a new random quotient of G, at least if we knew that G is torsion-free. This is not known (§ IV.f.), but the results of [Oll04] imply that if G_0 is of geometric dimension 2 then so is G. So in particular, taking a free group for G_0 and iterating Theorem 40 we get:

Proposition 43.
Let F_m be the free group on m generators $a_1, \ldots, a_m$. Let $(\ell_i)_{i\in\mathbb{N}}$ be a sequence of integers. Let $d < -\log_{2m} \rho(F_m)$ and, for each i, let R_i be a set of random words of length ℓ_i at density d as in Theorem 40.

Let $R = \bigcup R_i$ and let $G = F_m/\langle R\rangle$ be the (infinitely presented) random group so obtained.

Then, if the ℓ_i's grow fast enough, with probability arbitrarily close to 1 the group G is a direct limit of non-elementary hyperbolic groups, and in particular it is infinite.

It is not easy to follow the details of [Oll04, Oll05a, Ch93] closely enough to obtain an explicit necessary rate of growth for the ℓ_i's, although $\ell_{i+1} \geqslant \mathrm{Cst}.\ell_i$ is likely to work.

The techniques used in [Gro03] to get iterated quotients are different from those of [Oll05a] and of more geometric inspiration (see § III.2. or [Gro01a, Oll-c]); in particular, therein property (T) is used to gain uniform control on the critical densities of all successive quotients. The drawback is that these techniques only work for very small densities.

II.3. Counting one-relator groups

On a very different topic, consideration of generic-case rather than worst-case behavior for algorithmic problems in group theory (most notably the isomorphism problem) led I. Kapovich, Myasnikov, Schupp and Shpilrain, in a series of closely related papers [KSS, KS, KS05, KMSS05, KMSS03], to the conclusion that generic elements are often nicely behaved. The frontier between properties of one-relator groups and properties of a typical word in the free group is faint; for this review we selected an application where the emphasis is really put on the group, namely, evaluation of the number of distinct one-relator groups up to isomorphism.

The isomorphism problem for finite presentations is generally undecidable (see e.g. the very nice [Sti82] for an introduction to the word and isomorphism problem, or the end of chapter 12 of [Rot95]). It has been solved for the class of torsion-free hyperbolic groups with finite outer automorphism group ([Sel95], see also [Pau91]), which contains generic one-relator groups since their outer automorphism group is trivial [KSS]. However, having an algorithm for the isomorphism problem does not provide an estimate of the number of isomorphism classes.

For one-relator groups, the basic idea is as follows: Since generic relators are by definition much more numerous than particular relators, if we can show that one-relator groups with a generic enough relator are mutually non-isomorphic, then we will get a sharp estimate of the number of isomorphism classes of one-relator groups.

Let $I_\ell(m)$ be the number of isomorphism classes of one-relator group presentations $\langle\, a_1, \ldots, a_m \mid r \,\rangle$ with $|r| \leqslant \ell$. Of course $I_\ell(m)$ is less than the number of cyclically reduced words of length at most ℓ; this crude estimate can be improved since taking a cyclic permutation of r does not change the group. Now the number of cyclically reduced words of length $\leqslant \ell$ up to cyclic permutation is about $(2m-1)^\ell/\ell$. Moreover, some trivial symmetries (such as exchanging the generators $a_1, \ldots, a_m$ or taking inverses) decrease this estimate by some explicit factor depending only on m. Actually the estimate found this way is sharp [KS, KSS]:

Theorem 44.
The number $I_\ell(m)$ of isomorphism classes of one-relator groups on m generators, with the relator of length at most ℓ, satisfies

$$I_\ell(m) \sim \frac{1}{m!\, 2^{m+1}} \, \frac{(2m-1)^\ell}{\ell}$$

when $\ell \to \infty$.

Once Theorem 19 is known the result is relatively simple. Indeed, a theorem of Magnus ([LS77], Prop. II.5.8) implies that if two elements of the free group generate the same normal subgroups then they are the same up to conjugation and inversion. If two generic one-relator groups are isomorphic, then Theorem 19 implies that after applying some automorphism of the free group, the two relators have the same normal closure, and thus are essentially the same by Magnus' Theorem. So the isomorphism problem for generic one-relator groups reduces to the problem of knowing when an element in the free group is the image of another under some automorphism of the free group.

The orbits of the automorphism group of a free group are well studied and generated by so-called Whitehead moves ([LS77], I.4): especially, if two elements lie in the same orbit and are of minimal length within this orbit, they can be transformed into each other by means of non-length-increasing Whitehead moves. But for generic elements it can be easily shown that the action of the Whitehead moves increases length except for some trivial cases, so that generic elements do not lie in the same orbit.

In other words, a generic element cannot be "simplified" by action of automorphisms of the free group. This same minimal representation idea allowed to get an estimate [KS] of Delzant's T-invariant for generic one-relator group (this invariant, defined in [Del96b], attempts to measure the minimal possible "complexity" of presentations of a given group).

The only (but crucial) place above where one-relatorness plays a role is the use of the Magnus theorem, which has no known replacement for the case of several relators (even generic ones). On the other hand, genericity really lies at the core of the argument: the idea is that for counting matters, particular annoying cases can be discarded and only the nicest, typical cases can be treated.

Part III.
Applications: Random ingredients in specific constructions

What non-probabilists call “the probabilistic method” is the use of random constructions to prove existence theorems and to build new objects and (counter-)examples. Badly enough, this is often seen as the only possible justification for the introduction of random tools in a classical field.

Random groups fit into this scheme. Up to now, the main application of random groups is the construction by Gromov [Gro03] of a finitely presentable group whose Cayley graph (quasi-)contains an infinite family of expanding graphs and which contradicts the Baum-Connes conjecture with coefficients [HLS02]. We give the roadmap to this construction in § III.1. and § III.2.

A second application (§ III.3.), using the same tools together with a construction of Rips [Rip82], allowed in [OW-a] to construct Kazhdan groups whose outer automorphism group contains an arbitrary countable group, answering a question of Paulin (in the list of open problems in [HV89]). As was noted by Cornulier [Cor-b], this implies in particular that any discrete group with property (T) is a quotient of a torsion-free hyperbolic group with property (T). The technique is flexible and provides other examples of Kazhdan groups with prescribed properties.

Let us insist that groups constructed this way cannot pretend to “typicality”: in each case the random constructions are twisted in ways specific to the goal to achieve. The process of building a group containing a family of expanders starts with the choice of such a family and uses it to define the group; the expanders do not appear out of the blue in a plain random group (compare the techniques in [AC04], though: it may be that Cayley graphs of plain random groups contain lots of interesting families of graphs).

The common tool to both constructions above is Gromov’s powerful and flexible generalization of small cancellation theory to group presentations arising from labelled graphs. When everything goes well, the said graph embeds in the Cayley graph of the group, thus allowing “shaping” of Cayley graphs. Moreover, this extension of small cancellation is compatible with property (T), whereas usual small cancellation is not [Wis04], showing that a really new class of groups is accessible this way.

III.1. Shaping Cayley graphs: graphical presentations

Before we state the results, let us describe this graphical presentation tool. It is discussed in the last paragraph, “Random presentations of groups”, in [Gro00], and more thoroughly in sections 1 and 2 of [Gro03]. (The idea of representing subgroups of the free group by labelled graphs goes back to [Sta83], but therein the emphasis is on the subgroup and not on the quotient group “presented” by the graph.)

III.1.a. Labelled graphs and group presentations. Let us state some vocabulary geometrizing group presentations. Let $a_1, \ldots, a_m$ be our usual generators and let B be the following standard labelled graph: B consists of one single vertex and m oriented loops, univocally labelled with the generators $a_1, \ldots a_m$. Now a *word* in the $a_i^{\pm 1}$ is simply a path in the labelled graph B. A *reduced word* is an *immersed* path $P \looparrowright B$.

A *labelled graph* is a graph Γ together with a graph map $\Gamma \to B$, i.e. a graph in which every edge bears a generator a_i with an orientation. It is said to be *reduced* if this map is an immersion; this amounts to not having two distinct edges with identical labels originating (or ending) at the same vertex (this is called "folded" in [Sta83], but this terminology is less consistent with the case of reduced words, reduced van Kampen diagrams, etc.).

Gromov's idea is that to a labelled graph we can associate the group presentation whose relators are all the words read on cycles of the graph. More precisely, let Γ be a labelled graph and let x_0 be any basepoint in Γ. The labelling $\varphi : \Gamma \to B$ defines a map $\varphi : \pi_1(\Gamma, x_0) \to F_m$, sending a closed path to its label. The *group presented by* Γ is by definition the group

$$G = \langle\, a_1, \ldots, a_m \mid \Gamma \,\rangle = F_m / \langle \varphi(\pi_1(\Gamma), x_0) \rangle$$

(when Γ is not connected, this is defined as $G = F_m / \langle \cup\, \varphi(\pi_1(\Gamma, x_i)) \rangle$ taking a basepoint in each connected component). If $\pi_1(\Gamma)$ is generated by the cycles $(c_i)_{i \in I}$ then a cheaper presentation for G is

$$G = \langle\, a_1, \ldots, a_m \mid \varphi(c_i)_{i \in I} \,\rangle$$

and note that I can be taken finite if Γ is finite. Note also that changing the basepoint amounts to taking some conjugate of the image $\varphi(\pi_1(\Gamma))$, so that the group defined by Γ is unchanged. We will call $\langle\, a_1, \ldots, a_m \mid \Gamma \,\rangle$ a *graphical presentation* for G.

The group G is of course the fundamental group of the 2-complex obtained by gluing a disk in B along each of the paths $\varphi(c_i)$ where the c_i are the simple cycles of Γ.

Most importantly, when Γ consists of a disjoint union of circles, then we get back the usual notion of group presentation. The relators are cyclically reduced if and only if the labelling is reduced.

The Cayley graph $\mathrm{Cay}(G, (a_i))$ is itself a labelled graph. By definition of G, the label of any cycle in Γ is a relation in G and so can be read on some closed path in the Cayley graph. Consequently, if we fix a basepoint x_i in each connected component of Γ, and any basepoint $y \in \mathrm{Cay}(G, (a_i))$, then there is a unique label-preserving map

$$\varphi : \Gamma \to \mathrm{Cay}(G, (a_i))$$

sending each x_i to y, which we denote (again!) by φ since it commutes with the labelling maps to B.

Of course nothing guarantees that in general this map will be injective. It could happen that G is trivial, for example. But if φ is injective, then we have succeeded in embedding a graph in the Cayley graph of some group. This is what Gromov did with Γ a family of expanders, as we will describe in § III.2. For the moment, we turn to the description of a small cancellation condition ensuring this injectivity.

III.1.b. Graphical small cancellation. The central notion of small cancellation is that of *piece*: a piece is a word that can be read twice in the relators of a presentation. Here this notion generalizes as follows: Let $\Gamma \xrightarrow{\varphi} B$ be a labelled graph. A *piece* in Γ is a word $P \xrightarrow{\psi} B$ which can be read at two different places on Γ, that is, such that there are two distinct immersions $P \overset{i_1}{\looparrowright} \Gamma \xrightarrow{\varphi} B$ and $P \overset{i_2}{\looparrowright} \Gamma \xrightarrow{\varphi} B$ (preserving the labels, of course, i.e. $\varphi \circ i_1 = \varphi \circ i_2 = \psi$).

What matters for small cancellation is the size of pieces compared to the size of the relators on which they appear. Here the role of relators is played by cycles in the graph. So we define the *relative length* of a piece P to be the maximum of the ratio $|P| / |C|$ over all immersed cycles $C \looparrowright \Gamma$ such that P appears on C i.e. there exists $P \looparrowright C \looparrowright \Gamma$.

Definition 45.
A labelled graph Γ satisfies the graphical small cancellation condition *$Gr'(\alpha)$ if the relative length of any piece in Γ is less than α.*

It should be clear that when Γ is a disjoint union of circles, this reduces to the traditional $C'(\alpha)$ small cancellation condition. The well-known $C'(1/6)$ theory extends to the new framework. Similarly one could define the combinatorial $Gr(p)$ condition asking for no cycle in Γ to be the union of fewer than p pieces.

The following is a much simplified version of the statements in section 2.2 of [Gro03] (see § III.1.e. below for a more general setting). Gromov uses general geometric arguments, but the version presented here is easy to prove using the usual combinatorial techniques of small cancellation (see [Oll-a] or [Wis]).

Theorem 46.
Let Γ be a labelled graph which is spurless (i.e. with no valency-1 vertex) and reduced. Suppose that Γ satisfies the $Gr'(1/6)$ graphical small cancellation condition.

Then the group G defined by the graphical presentation Γ enjoys the following properties:

- *If Γ is finite, G is hyperbolic; if Γ is infinite G is a direct limit of hyperbolic groups.*
- *It is torsion-free, of geometric dimension 2, of Euler characteristic $1-m+b_1(\Gamma)$. In particular if $b_1(\Gamma) > m$ it is infinite and not free.*
- *The natural map of labelled graphs from Γ to the Cayley graph of G (for any basepoint choice) is an isometric embedding for the graph distances; in particular it is injective.*
- *The length of the shortest cycle in the Cayley graph is equal to that in Γ.*

More properties of usual small cancellation still hold, such as asphericity of the "standard" presentation and a kind of Greendlinger lemma (see details in [Oll-a]).

The reducedness assumption is necessary: otherwise we could add arbitrary long paths labelled by words which are trivial in the free group, thus artificially decreasing the relative length of pieces by increasing the length of cycles. Spurs do not change the group defined by the graph, but might not embed isometrically, unless $Gr'(1/8)$ holds.

The idea of one of the possible proofs is to consider van Kampen diagrams all faces of which bear a word read on some cycle of the graph, and are "minimal" in the sense that the word read on the boundary of the union of two adjacent faces is *not* read on a cycle of the graph (otherwise we merge the two faces). Such "minimal" diagrams (locally) satisfy the usual $C'(1/6)$ condition hence a linear isoperimetric inequality (w.r.t. to the set of relators made of all words read on cycles of the graph), and hyperbolicity follows by the remark that any cycle in the graph can be written as a concatenation of linearly many cycles of bounded sizes, so that we can replace this infinite presentation by a finite one while keeping a linear isoperimetric inequality. Asphericity, cohomological dimension and isometric embedding of the graph into the Cayley graph require a little more work. Although this basic idea is simple, there are some delicate topological details [Oll-a].

III.1.c. Random labellings are $Gr'(1/6)$. One of the interests of the Gr' condition is that random labellings of a graph satisfy it very probably. A random labelling of a graph is simply the choice, for each edge of the graph, of a generator a_i together with an orientation, picked at random among the $2m$ such possible choices. Generally this does not result in a reduced labelling, but we can reduce the graph by performing the necessary edge identifications (the "folding" of [Sta83]).

Of course, if there are "too many" cycles in the graph, the group will tend to be trivial. According to the spirit of the density model (Def. 7), this "too many" has to be defined with respect to the length of the cycles: the longer the cycles, the more of them we can tolerate. A way to achieve this is to subdivide the graph, i.e. replace each edge with 100 (say) consecutive edges. For a given unlabelled graph Γ, we will denote $\Gamma^{/j}$ its j-subdivision (each edge is replaced with j edges). Another interpretation of this is to use edge labels which are length-j words in the generators instead of single generators. Subdividing amounts to decreasing the "density" of the graph, which decreases the expected size of pieces (compare equation $(*)_\tau$ in section 1.2 of [Gro03] with a density). The same effect could be achieved by increasing the number m of generators.

We give here an oversimplified version of statements in paragraphs 1.1, 4.6 and 4.8 of [Gro03] (which deal, much more generally, with random quotients of hyperbolic groups by graphical presentations, see § III.1.e.). A proof of this particular case can be found in [OW-a]. We will need some "bounded geometry" assumptions on the graph, bounding the valency of vertices and the diameter/girth ratio. The *girth* of a graph is defined as the length of the shortest non-trivial cycle in it. It plays the role of the length of the relators in the density case.

Proposition 47.
For any $\alpha > 0$, for any number of generators $m \geqslant 2$, for any $v \in \mathbb{N}$ and $C \geqslant 1/2$ there exists an integer j_0 such that for any $j \geqslant j_0$ the following holds:

For any graph Γ satisfying the following conditions:

- *The valency of any vertex of Γ is at most v.*
- *The girth and diameter of Γ satisfy $\operatorname{Diam}\Gamma \leqslant C \operatorname{girth}\Gamma < \infty$.*

then a random labelling of the j-subdivision $\Gamma^{/j}$, once reduced, satisfies the $Gr'(\alpha)$ condition, with probability arbitrarily close to 1 if $\operatorname{girth}\Gamma$ is large enough (depending on α, v, C).

In particular for $\alpha < 1/6$ the conclusions of Theorem 46 hold.

Moreover, the metric distortion induced by the reduction step is controlled (see the last section of [Oll-a], or [OW-a]).

III.1.d. Random labellings of expanders entail Kazhdan's property (T). The group defined by a graphical presentation inherits some spectral properties of the graph, at least when the labelling is random. In particular, if the Laplacian on the graph has a large enough spectral gap, then the group defined by a random labelling will have property (T). Section 1.2 of [Gro03] mentions (generalizations of) this, and the whole section 3 of [Gro03] is devoted to building a general framework encompassing the usual spectral criteria for property (T) mentioned above in I.3.g. (see references there). We give only the following statement, a detailed proof of which was written by Silberman ([Sil03], Corollary 2.19):

Theorem 48.
Given $v \in \mathbb{N}$, $\lambda_0 > 0$ and an integer $j \geqslant 1$ there exists an explicit g_0 such that if Γ is a graph with $\operatorname{girth}\Gamma \geqslant g_0$, $\lambda_1(\Gamma) \geqslant \lambda_0$ and every vertex of which has valency between 3 and v, then a random labelling of the j-subdivision $\Gamma^{/j}$ defines a group with Kazhdan's property (T), with probability tending to 1 as the size of Γ tends to infinity.

The idea is as follows: Property (T) is related to the way the random walk operator acts on equivariant functions from the group to unitary representations of it. Now in the case of graphical presentations, by construction any equivariant function on the group (which is determined by its value at e) can be lifted to a "label-equivariant" function on the graph since cycles in the graph are labelled by relations in the group. If moreover the labelling was taken at random, then a random walk in the graph "simulates" a random walk in the group in the sense that the labels encountered by a random walk in the graph are plain random words (at short times). So if the graph has a large spectral gap, it is possible to transfer the spectral inequality to the random walk operator on the group. The details can be found in [Sil03].

Note that the first step (lifting equivariant functions) follows only from the definition of graphical presentations, whereas the second one uses the fact that the labelling was random (in some weak, statistically testable sense).

III.1.e. Generalizations: relative graphical presentations, and more. A labelled graph can also be used to define a quotient of an arbitrary marked group, by quotienting the group by the words read on cycles of the graph. This is a key step used by Gromov in the wild group construction described below (§ III.2.).

Just as ordinary small cancellation theory can be extended from quotients of the free group to quotients of a given hyperbolic group by elements satisfying a "relative small cancellation" condition ([Del96a], [Ch94], section 4 of [Ols93], section 5.5 of [Gro87]), an analogue to Theorem 46 holds when the initial group is hyperbolic (maybe with some restriction on torsion) instead of free.

In [Oll-c] an elementary version of Gromov's statements is given, which can be proven using the traditional van Kampen diagram approach of [Oll04], combined with the combinatorial arguments specific to the graphical case as in [Oll-a].

But Gromov proved this in a more general context using "rotation families of groups", where purely geometrical arguments can be given. The context is a group G acting properly and cocompactly by isometries on some hyperbolic space X; we want to study the quotient of G by a normal subgroup R.

In non-graphical small cancellation theory (relative to a hyperbolic group G), R is generated by elements (u_i) and all their conjugates; with each u_i is associated a geodesic U_i in X invariant under u_i; a conjugate of u_i will be associated with the corresponding translate of U_i. Small cancellation for the family (u_i) (relative to G) is equivalent to the family of all U_i's and their translates not to travel close to each other for a "too long" time (the time is measured w.r.t. the minimal displacement of the action of the u_i's on X, namely, less than 1/6 of this displacement; closeness is measured w.r.t. the hyperbolicity constant of X). For graphical small cancellation, say we have a connected labelled graph Γ; lift its universal cover $\tilde{\Gamma}$ to G and take the corresponding orbit U in X (this is a tree), together with all its translates (the translates correspond to conjugate lifts of $\tilde{\Gamma}$ to G); if this family of trees in X satisfies the same condition as above (not travelling close to each other for a long time), then the quotient of G by the labelled graph will be hyperbolic again. If Γ is not connected we get as many U_i's as there are connected components (plus their translates).

In section 2 of [Gro03], Gromov exposes a (difficult to read) general terminology and conditions for these ideas to work. Elements of proof are scattered in four papers (section 2 of [Gro03], sections 6–7 of [Gro01a], sections 25–32 of [Gro01b], section 10 of [Gro01c]). This framework seems to be quite powerful.

A simpler proof can be given in the case of *very* small cancellation (with 1/6 replaced by some tiny constant), using $\mathrm{CAT}(-1, \varepsilon)$ spaces. The idea of the proof, very neatly described at the beginning of [DG] (see also [Del-b] and Gromov's papers just cited) and fully developed later in that paper, is as follows: we have a group $G = \pi_1(X)$ acting properly cocompactly by isometries on a hyperbolic space $\tilde{X}$, and we want to quotient G by a normal subgroup R; the quotient is, of course, the fundamental group of the space X' obtained by gluing disks to X along loops in X corresponding to generators of the normal subgroup R. The idea is to endow these disks with a metric of constant negative curvature turning them into hyperbolic cones. This allows to check that X' is *locally* a $\mathrm{CAT}(-1, \varepsilon)$ space, and the Cartan-Hadamard theorem (or local-global principle for hyperbolic spaces) then allows to conclude that the universal cover $\widetilde{X'}$ is globally $\mathrm{CAT}(-1, \varepsilon)$, hence hyperbolicity of G/R.

This idea of metrizing Cayley complexes, applied in [DG] to the Burnside problem, looks very promising (see § IV.i.).

III.2. Cayley graphs with expanders

In [Gro03] (as announced in [Gro00]), Gromov constructs a finitely generated group whose Cayley graph "quasi-contains" a family of expanding graphs and which thus admits no uniform embedding into the Hilbert space. The main idea is to use a graphical presentation arising from a random labelling of these expanders.

Recall (see e.g. [Lub94], [DSV03]) that a family of expanding graphs (or expanders) is a sequence of graphs $(\Gamma_i)_{i\in\mathbb{N}}$ of bounded valency, of size tending to infinity, such that the first eigenvalue of the discrete Laplacian on them is bounded away from 0 when $i \to \infty$. A uniform embedding of metric spaces is a map φ such that there exists a function $f : \mathbb{R} \to \mathbb{R}$ with $\mathrm{dist}(\varphi(x), \varphi(y)) \geqslant f(\mathrm{dist}(x, y))$ and $f(x) \to \infty$ when $x \to \infty$.

One of the reasons for the interest in this paper is that, as proven by Higson, V. Lafforgue and Skandalis in [HLS02], this implies failure of the Baum-Connes conjecture with coefficients for this group. The initial stronger motivation was to refute the Novikov conjecture. Introducing these conjectures is beyond the scope of this paper and the author's field of competence. We refer the reader to [KL05, Val02b, Ska99, Hig98]. Gromov's group is a direct limit of hyperbolic groups; for hyperbolic groups, the Novikov conjecture [CM88, CM90], existence of a uniform embedding into the Hilbert space [Sel92] and the Baum-Connes conjecture [Laf02, MY02] hold. For the link between those last two properties see [Yu00, STY02].

Theorem 49.
For any $\varepsilon > 0$ there exists a finitely generated, recursively presented group G, a family of expanders $(\Gamma_i)_{i\in\mathbb{N}}$, constants $A, B > 0$ and maps φ_i sending the vertices of Γ_i to vertices of $\mathrm{Cay}(G)$ such that

$$A\,(\mathrm{dist}(x, y) - \varepsilon\,\mathrm{Diam}\,\Gamma_i) \leqslant \mathrm{dist}(\varphi_i(x), \varphi_i(y)) \leqslant B\,\mathrm{dist}(x, y)$$

for any i and $x, y \in \Gamma_i$, where the distance in Γ_i is the ordinary graph distance and the distance in $\mathrm{Cay}(G)$ is w.r.t. some fixed finite generating set.

Consequently there exists a finitely presented group admitting no uniform embedding into the Hilbert space.

All the ingredients of the proof can be found in [Gro03], though lots of details are omitted. Gromov apologizes in the introduction that he chose not to write "a few technical lemmas, with a straightforward half-page proof each" but rather to "uncover the proper context rendering [...] the proofs tautological", and then adds "A reader may find it amusing to play the game backwards by reducing the present paper to seven pages of formal statements and proofs". This is still waiting to be done, though some parts of the job are written [DG, Oll-c, Oll-a, Sil03]. Full understanding and exploitation of these "contexts" will doubtlessly be an important source of new results and techniques.

The principle of the proof is as follows (and the technical conditions needed for it to work are stated below in Definition 50 and Theorem 51, extracted from [Oll-c]): Start with the free group F_2 and any family of expanders Γ_i. Put a random labelling on some subdivision $\Gamma_1^{/j}$ of Γ_1 and let G_1 be the group given by the graphical presentation Γ_1. According to Proposition 47 and Theorem 46, if j is large enough, G_1 will be a non-trivial hyperbolic group. As described in § III.1.a., there will be a natural graph map $\Gamma_1^{/j} \to \mathrm{Cay}(G_1)$, which is actually a quasi-isometric embedding.

Then consider a random labelling of a subdivision of Γ_2 and let G_2 be the quotient of G_1 by the graphical presentation Γ_2. Applying the "relative" version of proposition 47, as described in § III.1.e., G_2 will be a non-trivial hyperbolic group, provided the girth of Γ_2 is large compared to the hyperbolicity constant of G_1 (which is of the same order of magnitude as $\mathrm{Diam}\,\Gamma_1$). Up to taking a subsequence of the family of expanders we can always suppose that $\mathrm{girth}\,\Gamma_i \gg \mathrm{Diam}\,\Gamma_{i-1}$, which allows to define inductively a hyperbolic group G_i obtained by quotienting G_{i-1} by a random labelling of (a subdivision of) Γ_i. The group G_i comes with a natural graph map from $\Gamma_i^{/j}$ to its Cayley graph.

The group G is then obtained as the direct limit of all G_i's. It is *not* finitely presented, but can be recursively presented by replacing randomness by pseudo-randomness. Indeed, the graphical small cancellation property used here is algorithmically checkable (even relatively to a given hyperbolic group!), so that we can write a program enumerating all labellings of Γ_1, testing whether they are $Gr'(1/6)$, stopping at the first such labelling found (which exists by the randomness argument), then outputting a presentation for G_1; enumerating all labellings of Γ_2 and testing whether they are in small cancellation relative to the explicit hyperbolic group G_1, outputting the first such labelling of Γ_2, etc. Note that this requires to have a *recursive* construction for the expanders Γ_i too.

This provides a recursive enumeration of the presentation of the limit group G. Then applying Higman's embedding theorem (Theorem 12.18 in [Rot95], Theorem IV.7.3 in [LS77]) provides a *finitely presented* group H in which G embeds. Note that an embedding of a finitely presented group is always a uniform embedding (since there are only finitely many elements in balls of the image of the initial group), so that H does not uniformly embed into the Hilbert space if G does not.

The subdivision step amounts to label each edge of Γ_i with a random word of length j rather than with a single generator. This allows to reduce "density" of the graphical presentation, by increasing the relator length (measured by the girth) without changing the number of relators. It is very important to use the same j for *all* the Γ_i's: indeed we only get a graph map from $\Gamma_i^{/j}$ to the Cayley graph of G, which of course induces a map from the vertices of Γ_i to $\mathrm{Cay}(G)$ with a j times larger Lipschitz constant, so that if j goes uncontrolled then so do the metric properties of the embedding. In other words, a bounded subdivision of a family of expanders is still a family of expanders but this is false for unbounded subdivisions.

Another important point is that the "critical density" for non-triviality of random quotients of the G_i's could decrease to 0 when $i \to \infty$, thus resulting in groups that are more and more reluctant to adding new relations (forcing to increase j). As results from Theorem 40, this critical density is controlled by the spectral radius of the random walk on G_i. So it is important to get a uniform control on this spectral radius for all G_i's. Actually property (T) of a group entails such a uniform control of the spectral radii of all of its quotients. So if G_1 has property (T) we are done, and this results from Theorem 48. (Another way to proceed is to replace the initial group F_2 with a hyperbolic group having property (T). Yet another, maybe most natural way is to use Theorem 42 which states that the spectral radius is almost unaffected by random quotients.)

The "density" of a graphical presentation is not only controlled by the size of cycles in the graph but also of course by the number of cycles. Demanding that these graphs have bounded geometry (valency, diameter/girth ratio) ensures that density remains bounded.

So, putting all the constraints altogether, we get the following conditions for the construction to work (see [Oll-c]). Note that having a family of expanders is *not* required for the process of building the limit group, so that it is possible to get Cayley graphs containing other interesting families of graphs.

Definition 50.
A sequence $(\Gamma_i)_{i\in\mathbb{N}}$ of finite connected graphs is good for random quotients *if there exists $v \geqslant 1$ and $C, C' \geqslant 1$ such that for all i we have:*

- $\operatorname{girth}\Gamma_i \to \infty$;
- $\operatorname{Diam}\Gamma_i \leqslant C \operatorname{girth}\Gamma_i$;
- *For any $x \in \Gamma_i, r \in \mathbb{N}$, the ball $B(x,r)$ of radius r in Γ_i satisfies $\#B(x,r) \leqslant C'\, v^r$.*

Theorem 51.
Let $(\Gamma_i)_{i\in\mathbb{N}}$ be a sequence of finite connected graphs which is good for random quotients.

Then for any $\varepsilon > 0$ there exists a finitely generated group G_∞, an increasing sequence i_k of integers, an integer $j \geqslant 1$ and a constant $A > 0$ such that, for any $k \in \mathbb{N}$, there exists a map of graphs $\varphi_k : \Gamma_{i_k}^{/j} \to \operatorname{Cay}(G_\infty)$ from the j-subdivision of Γ_{i_k} to the Cayley graph of G_∞, which is quasi-isometric in the following sense:

For any $x, y \in \Gamma_{i_k}^{/j}$ we have

$$A\left(\operatorname{dist}(x,y) - \varepsilon \operatorname{Diam}\Gamma_{i_k}^{/j}\right) \leqslant \operatorname{dist}\left(\varphi_k(x), \varphi_k(y)\right) \leqslant \operatorname{dist}(x,y)$$

where the distance in $\Gamma_{i_k}^{/j}$ is the usual graph distance and the distance in $\operatorname{Cay}(G_\infty)$ is that w.r.t. a fixed finite generating set.

The group G_∞ has some labelling of the union of the $\Gamma_{i_k}^{/j}$'s as a graphical presentation. This presentation is aspherical (in the sense given in [Oll-a]) and turns G_∞ into a direct limit of hyperbolic groups of geometric dimension 2.

Finally, if the family of graphs Γ_i is recursive, then this graphical presentation can be assumed to be recursive.

Note that the size of the fibers $\varphi_k^{-1}(x)$ is bounded by $jv^{\varepsilon \operatorname{Diam} \Gamma_{i_k}}$, so that if $\#\Gamma_{i_k}$ grows reasonably fast (as is the case for expanders), then the small fiber condition appearing in [HLS02] is satisfied.

Theorem 49 now follows from the above and the existence of a recursive family of expanders (e.g. Theorems 7.4.3 and 7.4.12 of [Lub94], or [DSV03]):

Theorem 52.
There exists a recursively enumerable family of graphs $(\Gamma_i)_{i\in\mathbb{N}}$ such that:

- $\#\Gamma_i \to \infty$;
- $\inf_i \lambda_1(\Gamma_i) > 0$ *(the Γ_i are expanders);*
- *for all i, Γ_i is regular of valency v;*
- *there exist C_1, C_2, C_3 such that* $\log \#\Gamma_i \leqslant C_1 \operatorname{Diam} \Gamma_i \leqslant C_2 \operatorname{girth} \Gamma_i \leqslant C_3 \log \#\Gamma_i$ *for all i.*

Besides [Gro03], more information on Gromov's construction can be found in [Ghy03, Pan03, Oll-c]. Useful elements of proof appear in [DG, Oll-a, Del-b, Oll-c, Sil03] and of course in [Gro03, Gro01a, Gro01b, Gro01c]. The link with the Baum-Connes conjecture is proven in [HLS02].

III.3. Kazhdan small cancellation groups?

A more modest application of Gromov's random graphical presentations is that they allow a nice mixture of small cancellation properties and property (T), using Proposition 47 together with Theorem 48. This contrasts with ordinary $C'(1/6)$ groups, which do not have property (T) (unless finite) by a result of Wise (Corollary 1.3 in [Wis04]).

This allows the construction of Kazhdan groups with somewhat unexpected properties, using the flexibility of small cancellation groups. The main tool here is a short exact sequence coined by Rips [Rip82]. Namely, for every countable group Q, Rips constructed an exact sequence $1 \to N \to G \to Q \to 1$ where G is a $C'(1/6)$ group and the kernel N is finitely generated. Pathologies of Q often lift to G in some way. Carefully adding a random graphical presentation to G adorns N with property (T), namely [OW-a]:

Theorem 53.
For each countable group Q, there is a short exact sequence $1 \to N \to G \to Q \to 1$ such that G has a $Gr'(1/6)$ presentation and N has property (T). Moreover, G is finitely generated if Q is, and G is finitely presented (hence hyperbolic) if Q is.

As noted by Cornulier [Cor-b], this easily implies a kind of universal property for hyperbolic Kazhdan groups:

Corollary 54.
Every countable group with property (T) is a quotient of a $Gr'(1/6)$ hyperbolic group with property (T).

Indeed, in the exact sequence above, if both Q and N have (T) then G has (T) [HV89]. If Q is finitely presented, then so is G and thus G is hyperbolic. If Q is not finitely presented, a theorem of Shalom (Theorem 6.7 in [Sha00]) provides a finitely presented Kazhdan group of which Q is a quotient.

Another consequence of Theorem 53 is the following. Paulin asked (open problem 5 at the end of [HV89]) if a Kazhdan group can have an infinite outer automorphism group (this is impossible for a hyperbolic group by a result of [Pau91]). Actually this can happen and more precisely [OW-a]:

Theorem 55.
Every countable group embeds in the outer automorphism group of some Kazhdan group.

Indeed in the exact sequence above, Q acts on N by conjugation and this action happens not to be inner. In particular for finitely presented Q, the group N appears as a subgroup of some hyperbolic group.

Very different examples of Kazhdan groups with infinite outer automorphism groups were independently constructed by Cornulier [Cor-a] (as linear groups) and later by Belegradek and Szczepański [BSz] using relatively hyperbolic groups. Moreover Cornulier's example is finitely presented, thus answering positively a question in [OW-a].

Using the techniques in [BW05], it may be possible to show that actually every countable group is isomorphic to the outer automorphism group of some Kazhdan group. For finitely presented groups, this is shown in [BSz] up to finite index.

The main interest of the combined Rips sequence/random graphical presentation method is its flexibility. Using standard techniques it is straightforward to construct new groups with prescribed properties. In [OW-a] two easy examples are given. Recall a group G is called *Hopfian* if every surjective homomorphism $G \to G$ is injective, and *co-Hopfian* if every injective homomorphism $G \to G$ is surjective.

Theorem 56.
There exists a Kazhdan group which is not Hopfian, arising as a finitely generated subgroup of a $Gr'(1/6)$ infinitely presented group. There exists a Kazhdan group which is not co-Hopfian, arising as a finitely generated subgroup of a $Gr'(1/6)$ hyperbolic group.

For comparison, for hyperbolic groups the situation is as follows: Sela proved [Sel99] that every torsion-free hyperbolic group is Hopfian, and this was extended [Bum04] to any finitely generated subgroup of a torsion-free hyperbolic group (showing that the infiniteness of the presentation in the theorem above cannot be removed). Sela again (final theorem of [Sel97]) proved that a non-elementary torsion-free hyperbolic group is co-Hopfian if and only if it is freely indecomposable; hence, every Kazhdan hyperbolic group is co-Hopfian.

Once more, subsequent examples using different techniques are described in [Cor-a] and [BSz]. Noticeably, Cornulier's example of a Kazhdan non-Hopfian group (arising from a p-arithmetic lattice) is finitely presented.

We have attempted to demonstrate that random groups already produced some interesting new examples of groups. The techniques involved are flexible enough and hopefully more is to come.

Part IV.
Open problems and perspectives

I feel, random groups altogether may grow up
as healthy as random graphs, for example.
M. Gromov, *Spaces and questions*

The problems presented hereafter are varied in style and difficulty. Some of them amount to a cleaning of results implicit in the literature, others are well-defined questions, whereas the worst of them are closer to babbling on an emerging notion. Some are directly extracted from the excellent exposition of Gromov in the final chapter of [Gro93], and still unsolved.

Only problems directly pertaining to random groups are presented here. It must be stressed that Gromov's paper [Gro03] contains a lot of new, challenging ideas inspired by his random group construction but belonging to neighboring fields, which unfortunately could not be discussed here.

Disclaimer. The list of problems is provided "as is", without any warranty, either express or implied, including, but not limited to, the warranty of correctness, of interest, of fitness to any particular purpose (such as an article or thesis), or of non-triviality. We wish the reader good luck.

IV.a. What happens at the critical density? The usual question after a talk on random groups...

Asking whether a random group at density $d = 1/2$ is infinite or trivial might not be the right way of looking at things. The most promising and intriguing approach is to define a limit object for $\ell \to \infty$ and for definite $d < 1/2$, and then let $d \to 1/2$. The limit object would be as follows: by Theorem 13 the Cayley graph of the random group G is a tree up to distance $\ell(1-2d)$, and moreover the hyperbolicity constant is at most $4\ell/(1-2d)$. So it is natural to consider the metric space $\frac{1}{\ell}\mathrm{Cay}(G)$ where $\frac{1}{\ell}$ means we rescale the distance by this factor: this yields, for any ℓ, a $4/(1-2d)$-hyperbolic space which is a tree up to distance $1-2d$. It seems likely that for fixed d, for $\ell \to \infty$ this metric space converges *à la* Gromov-Paulin to some (maybe deterministic in some sense) non-locally compact metric space locally modeled on a real tree. This object would depend on d and be $4/(1-2d)$-hyperbolic. Letting then $d \to 1/2$ might bring a non-trivial object, maybe with some self-similarity or universality properties.

Another approach consists in letting simultaneously $\ell \to \infty$ and $d \to 1/2$. Indeed, Theorem 13 shows that the ball of radius $\ell(1-2d)$ in the Cayley graph is a tree. One is thus tempted to let $d \to 1/2$ and set $\ell = K/(1-2d)$ so that the length of the smallest relation in the group is kept constant (but the big problem is that this ℓ may be too small for Theorem 11 to hold). It may happen that for large enough values of K, the group converges (in law) to some non-trivial group with radius of injectivity

K, maybe infinitely presented. It may also happen that the group is trivial no matter how large K is.

If one sticks to the question of what happens when we take exactly $d = 1/2$ in the definition of the density model, one should note the following. If in the density model we take not $(2m-1)^{d\ell}$ but $P(\ell)(2m-1)^{d\ell}$ relators with P a subexponential term, for $d \neq 1/2$ this does not change the theorem. But for $d = 1/2$ triviality or infiniteness may depend on the subexponential term P. It might depend moreover on the details of the model (such as taking relators on the sphere or in the ball). Exact determination of these parameters might not be very relevant. Anyway, as a short answer, for $d = 1/2$ and $P(\ell) = 1$, it is easy to check (using the probabilistic pigeon-hole principle as in the comments after Theorem 11) that the random group has a positive probability (something like $(1/e)^{2m}$) to be trivial.

Some expect, however, that "all classical groups lie at $d = 1/2$" (using precise enough asymptotics for the $P(\ell)$ above?). By the way, note that property (T) holds at $d > 1/3$ and in particular at $d = 1/2$.

IV.b. Different groups at different densities? Another question, lying at the core of the density model, is to know whether density really has an impact on the random group.

The question is not exactly to know whether two random groups are mutually isomorphic or not: indeed two successive random samplings of a group at the same length and density will likely be non-isomorphic (although a proof of this would be very interesting and difficult, compare § II.3. for the one-relator case; see also [Gro93], p. 279). Rather one would like to know if the probability measures for distinct values of d become more and more different as $\ell \to \infty$. More precisely, one would like to know if, for every density d_0 and $\varepsilon > 0$, there exists a property of groups $P_{d_0,\varepsilon}$ which occurs with probability tending to 1 at density $d = d_0$, and with probability tending to 0 at any density $d \notin (d_0 - \varepsilon, d_0 + \varepsilon)$, as $\ell \to \infty$.

Since a random group at density d and length ℓ has Euler characteristic $1 - m + (2m-1)^{d\ell}$, for fixed m the number $d\ell$ can be recovered from the algebraic structure of the group. So it would be enough to recover any other combination of d and ℓ to get the answer. It is clear that *as marked groups*, with their standard generating set being known, random groups are different: indeed, for example the optimal isoperimetry constant $1-2d$ is provided by Theorem 13. But changing the generating set is a mess.

A very interesting but apparently difficult approach is suggested by Gromov in [Gro93] (p. 279). Let $G = F_m/N$ be a finitely presented group and define the *density* of this presentation as follows: there is an integer ℓ such that the normal closure of $N \cap B_\ell$ is N (where B_ℓ is the ball of radius ℓ in F_m); for $1 \leqslant k \leqslant \ell$ let the density $d_k = \log \#(N \cap B_k)/\log \#B_k$ (one can use spheres instead of balls) and let $d(F_m \to G) = \sup_{k \leqslant \ell} d_k$. Now let $d(G)$ be the infimum of these densities $d(F_m \to G)$ over all finite generating sets of G. The question is whether for a random group at density d this gives back d. Computation of densities of classical groups would also be interesting.

Gromov gives several other approaches in [Gro93], 9.B.(i).

The nicest thing would be to find group invariants depending continuously on density. The rank of the group is a discrete invariant varying with density, but is far from well understood (see § IV.d.). Pansu suggested the use of ℓ^p-cohomology, where the critical p might vary with d (see [Gro93], 9.B.(i) on $\ell_p H^1 \neq \{0\}$), or the conformal dimension of the boundary, but this approach has not yet been developed. It is wise to keep in mind the non-variation of the spectral gap (Theorem 26).

Note that this would not contradict Theorem 34 since, first, we do not expect independently picked random groups at the same density to be isomorphic, and second, random groups are not at all dense in $\mathcal{G}_m$.

IV.c. To (T) or not to (T). Property (T) for random groups is known to hold at density $> 1/3$ (Theorem 27), and not to hold at density $< 1/5$ (Theorem 32). There necessarily exists a critical density for property (T), since this property is inherited by quotients (indeed: if at a density d_0, property (T) occurs with positive probability, then at densities $d > d_0$ we can write the group presentation as a union of a large number of presentations at density d_0, and one of them is enough to bring property (T)). Determination of this critical density is a frustrating question.

The gap between the Haagerup property at $d < 1/6$ (Theorem 33) and failure of property (T) at $d < 1/5$ is probably just a technical weakness in [OW-b]. It would be very interesting to know whether, for random groups, property (T) starts just where the Haagerup property stops, so that these two properties, though not opposite, would be "generically opposite" (a possibility some people consider would be "sad").

Another question is whether property (T) holds at $d > 1/3$ for a random quotient (this is already asked in [Gro87], 4.5.C). It is trivially the case for quotients by random words, for any initial group (since property (T) is inherited by quotients) but in this case the random quotient might already be trivial at $d = 1/3$ (see Theorem 40) and so this statement could be empty. But it is very reasonable to expect the same holds for random quotients by elements in balls of hyperbolic groups as in Theorem 38: applying the criterion of [Żuk03] to the generating set made of all elements of length $\ell/3$ looks promising.

In the meantime, it is a good exercise to write a precise proof of the fact that random groups at density $d > 1/3$ have property (T) (in the density model, not in the triangular model, see discussion in § I.3.g.).

IV.d. Rank and boundary. Even such a simple invariant of groups as the rank (minimal number of elements in a generating set) is not known for random groups (except of course for random quotients of F_2). The rank does vary with density: by Theorem 17, at density 0 (and likely at small enough densities) it is equal to m, but it is easy to show that at density $d > 1 - \log(2k-1)/\log(2m-1)$, the rank is at most k (this follows from evaluating the probability that some relator can be written as a product of one generator followed by $\ell - 1$ generators chosen among k). This bound is very crude and probably not optimal (one may expect, for example, the rank to be 2 when $d \to 1/2$).

The rank would provide a non-continuous but nevertheless interesting invariant to prove that different densities produce different groups.

A possible approach is to generalize the method of Arzhantseva and Ol'shanskiĭ at density 0, which uses representation of subgroups by graphs and subsequent study of the exponential growth rate of the number of words which are readable on the graph. Combination of this with large deviation techniques for finite-state Markov chains may lead to a sharp estimate of the various exponents (densities) at play.

These techniques may be useful for other questions related to the algebraic structure. For example, Gromov asks ([Gro93], 9.B.(i)) whether a random group contains a non-free infinite subgroup of infinite index. Also, one would like to extend Theorem 44 on the number of one-relator groups to more relators (as asked by Gromov in [Gro93], p. 279), which, besides the interest of counting groups, would have implications for the problems discussed in § IV.b.

It is very likely that a random group is one-ended and in particular does not split as a free product (this is true at small densities and at $d > 1/3$, see § I.3.d.). Another question is unicity of the generating set up to Nielsen moves (compare Theorem 19); e.g. for $m = 2$ it can be shown that, for $d \geqslant 0.301\ldots$ (an apparently transcendental value coming from large deviation theory of the 4-state Markov chain describing reduced words in two letters), the pair (a_1^2, a_2) generates the random group, but it is not clear whether or not this pair is Nielsen-equivalent *in the group* to the standard generating pair (a_1, a_2).

IV.e. More properties of random groups. Any question which is meaningful for torsion-free hyperbolic groups may be asked for random groups. Some may even be answered.

Paragraph 1.9 of [Gro03] lists a few invariants "where a satisfactory answer seems possible": geometry of the boundary, L_p-cohomology, simplicial norm on cohomology, existence/non-existence of free subgroups, non-embeddability of random groups to each other, "something C^*-algebraic".

Another frequently asked question is the existence of non-trivial finite quotients of a random group and of residual finiteness. For any finite group H fixed in advance, it is easy to show that a random group with large enough defining relators will not map onto H. Exchanging the limits would provide hyperbolic groups without finite quotients. (See also the temperature model in § IV.k. below.)

IV.f. The world of random quotients. The theory of random quotients of given groups, the basic idea of which is that typical elements in a given group are the most nicely behaved, is at its very beginning. (Following Erdős, this also plays an increasingly important role in the—especially algorithmic—theory of finite groups, a subject we could not even skim over in this survey, see e.g. [Dix02].)

Theorem 38 and 40 only deal, for the moment, with quotients of torsion-free hyperbolic groups (which is nevertheless a generic class!). Of course there is no hope to extend these theorems to any initial group, if only because there exist infinite simple groups (but note that the "triviality" parts of these theorems extend to any group of exponential growth).

Nevertheless, the critical density 1/2 as in Theorem 38 seems to be quite a general phenomenon. Within a hyperbolic group, the density 1/2 principle might apply to random quotients by elements chosen in much more general subsets than the balls w.r.t. some generating set: more or less any large subset X not resembling too much to a line should do, i.e. quotienting by less than $\sqrt{\#X}$ elements randomly chosen in X should preserve hyperbolicity. The axioms defined in [Oll04] may help for this. This would have the advantage of decreasing the role of the generating set.

Identifying families of non-hyperbolic groups for which density 1/2 is critical would be very nice too.

The theory of random quotients works well for torsion-free hyperbolic groups. In the case of "harmful" torsion more complex phenomena occur (§ II.2.b., [Oll05b]). Identifying necessary and sufficient conditions on torsion (probably having to do with growth/cogrowth of the centralizers of torsion elements) for the random quotient theorems to hold, and identifying the kind of new phenomena (such as more than two phases [Oll05b]) which can happen in the presence of harmful torsion, would be interesting.

Speaking of torsion, it is not even clear whether a random quotient of a torsion-free hyperbolic group is still torsion-free. It is true however that geometric dimension 2 (which implies torsion-freeness) is preserved—this is what we use in every iterated quotient construction, as in Proposition 43 and Theorems 49 and 51 for the group with expanders. Using higher-dimensional complexes instead of the Cayley 2-complex (such as the Rips complex) or the techniques in [Del-a], may help get the result.

Theorem 39 states that the growth exponent is preserved when quotienting by random elements in a ball, and Theorem 42 states that the spectral radius is preserved when quotienting by random words; but it is likely that both are preserved whatever the model of random quotient. Elements to prove this appear in [Oll-e].

The methods used in [Oll04] to prove the phase transition theorems for random quotients of hyperbolic groups are partly geometric, partly combinatorial. On the other hand, those in [Gro03] and [DG] are almost purely geometric, but they do not allow to make it to the critical density and only work for "very" small cancellation [Del-b]. Even the basic density 1/2 theorem (Theorem 11) has a much more combinatorial than geometric proof. Geometrizing these proofs is a good challenge.

IV.g. Dynamics on the space of marked groups. The bad behavior of the isomorphism relation on the space of marked groups [Ch00] from the measurable point of view suggests an ergodic approach (part 4 in [Ghy03], 9.B.(g) in [Gro93]). The dynamics here comes from the action of the Nielsen moves on $\mathcal{G}_m$ (more precisely, the Nielsen moves on $2m$-tuples generate the isomorphism relation on $\mathcal{G}_m$ by a theorem of Tietze, see part 3 of [Ch00]). It seems likely, but is not known, that there is no non-trivial Borel measure on $\mathcal{G}_m$ invariant under this action. It would be nice, and perhaps important, to have an at least quasi-invariant measure.

There is a quite natural (family of) probability measure(s) on the space of all presentations of m-generated groups, coming from the temperature model (see § IV.k. below), which depends on a continuous parameter. This measure projects to a measure on $\mathcal{G}_m$, the properties of which (especially its behavior under Nielsen moves) must certainly be studied.

Besides, the study of continuity/measurability/average/whatever of the usual invariants of groups or presentations on $\mathcal{G}_m$ is interesting, as suggested in [Gro93], 9.B.(g).

Ghys noted that the complexity of $\mathcal{G}_m$ comes from lack of rigidity of the free group, and suggests that studying the space of quotients of a given marked hyperbolic group would keep all the nice properties of quotients of the free group (small cancellation, random quotients...), while maybe providing enough rigidity to allow better topological and measurable behavior, if the hyperbolic group has few automorphisms. This is of course related to § IV.f. above.

IV.h. Isoperimetry and two would-be classes of groups. Two natural properties related to isoperimetric inequalities of van Kampen diagrams arise naturally in random groups (including Gromov's group with expanders) and should be studied for themselves, independently of any probabilistic context.

The first one is stronger than mere hyperbolicity and generalizes small cancellation. Random groups, just as small cancellation groups, have the property that *any* (reduced) van Kampen diagram satisfies a linear isoperimetric inequality—whereas the definition of hyperbolicity asks that only one van Kampen diagram per boundary word satisfies such an inequality. This property implies, in particular, geometric dimension 2. But it is not stable by quasi-isometry, since for example taking the Cartesian product with a finite group introduces (a finite number of) spherical diagrams.

An interesting question is whether this property can be "geometrized", i.e. to modify this property so that it becomes invariant under quasi-isometry; a way to do this may be to ask that all van Kampen diagrams, after some *local* modifications, satisfy the isoperimetric inequality. This geometrized property might be equivalent, for example, to having a boundary of dimension one.

This certainly has to do with the "unfolded hyperbolicity" of [Gro01c], which is Gromov's newly coined name for the "local hyperbolicity" of section 6.8.U in [Gro87] (which asks that any "locally minimal" diagram satisfies the isoperimetric inequality; a link is explained with non-existence of conformal maps, and with any surface in the space having negative Euler characteristic); these considerations probably deserve more attention.

Groups in this class may keep lots of interesting properties of small cancellation groups.

The second property is the "homogeneous isoperimetric inequality". The usual way to write the isoperimetric inequality for a van Kampen diagram D is $|\partial D| \geqslant C\,|D|$ where $|D|$ is the number of faces of D. But a more natural way is a linear isoperimetric inequality between the boundary length of D and the sum of the lengths of the boundary paths of faces of D:

$$|\partial D| \geqslant C \sum_{f \text{ face of } D} |\partial f|$$

which is more homogeneous since it compares a length to a length, not a length to a number. For a finite presentation the two formulations are clearly equivalent (with a loss in the constant equal to the maximal length of a relator in the presentation).

This inequality is especially useful when facing a group presentation with relators of very different lengths, and is relevant also for infinite presentations. It naturally appears in $C'(\alpha)$ small cancellation theory (with the constant $C = 1 - 6\alpha$), in random groups (with $C = 1 - 2d$, see Theorem 13), in the few-relator model of random groups with various lengths (theta-condition of [Ols92]), in Champetier's work on cogrowth [Ch93], in computation of the hyperbolicity constant [Oll-b], in random quotients of hyperbolic groups (section 6.2 of [Oll04]), in iterated quotients (it is satisfied with $C = 1 - 2d$ under the assumptions of Proposition 43) and, noticeably, it is satisfied by the infinitely presented groups containing expanders constructed in [Gro03]. Maybe importantly, it allows a formulation of the local-global principle without loss in the constants and so seems to be the right assumption for it (Theorem 60 below, [Oll-f]).

So in lots of important contexts, even for finite presentations, this is the right way to write the isoperimetric inequality.

The main question is whether this has some intrinsic and/or interesting meaning for infinitely presented groups ("fractal hyperbolicity"? compare [Gro03], 1.7). The behavior under a change of presentation is unclear: for example the property is trivially satisfied (even for a finitely presented group) if the presentation consists of all relations holding in the group. One should probably restrict oneself to presentations with some minimality assumptions (e.g. the one which, given a set of generators, consists in beginning with an empty set of relators, and successively adding all relations which are not consequences of already taken, shorter relations), and study how the property is affected by elementary changes of the presentation.

IV.i. Metrizing Cayley graphs, generalized small cancellation and "rotation families". The generalized small cancellation theory used in [Gro03] (briefly described in § III.1.e. above) is developed in several papers [Gro03, Gro01a, Gro01b] (see also [DG, Del-b]). A single consolidated proof of the statements in section 2 of [Gro03] combined with a few examples (such as relative graphical small cancellation as stated in [Oll-c]) would be very useful. Compared to the traditional study of van Kampen diagrams, here the emphasis is put on geometric objects (such as lines for traditional small cancellation or trees for the graphical case) lying in a hyperbolic space acted upon by a group, and on conditions under which the space can be quotiented along these objects.

The approach can be purely geometric (as in [DG]) or in great part combinatorial (as in [Oll-a, Oll-c]). The geometric approach as written in [DG] does not work up to the optimal cancellation coefficient 1/6 but only for "very small" cancellation. But its strength is that, contrary to relative small cancellation, it can deal with quotients of a hyperbolic group by relators of length equal to the characteristic length of the group *plus* some large constant, whereas relative small cancellation needs the relation length to be a large constant *times* the characteristic length of the ambient hyperbolic group. This is why this approach succeeds in the case of the Burnside group.

The main idea is to put a non-trivial, negatively curved metric on the faces of the Cayley complex. This might be a step just as important as the jump from combinatorics of words to study of Cayley graphs and van Kampen diagrams. It is advocated in [DG] that hyperbolic groups can be made "much more hyperbolic" this way, in some intrinsic sense, than when just using the edge metric on the Cayley graph (or the Euclidean metric on the Cayley 2-complex). This technique is very flexible and may find many applications. At the very least it should provide a nice framework to re-interpret some classical results of hyperbolic group theory in.

IV.j. Better Cayley graphs with expanders? The construction of a Cayley graph with expanders may be simplified. The most direct way would be to find an explicit $Gr'(1/6)$ labelling of the whole family of expanders; this would provide both a shorter proof and an isometric embedding of the expanders, instead of quasi-isometric. Getting an *injective* (on the vertices of the expanders) quasi-isometric embedding would already be nice.

Another "flaw" of the construction is the final step using the Higman embedding theorem in order to get a finitely presented group: this keeps non-uniform embeddability into the Hilbert space, but the quasi-isometricity of the embedding of the expanders into the Cayley graph is lost, as are geometric dimension 2 and property (T), so the question of expanders quasi-isometrically contained in the Cayley graph of a finitely presented, maybe also Kazhdan group of dimension 2 is still open.

IV.k. The temperature model and local-global principles. Certainly one of the most important theoretical problems related to random groups.

All the random groups defined so far define a notion of asymptotically typical properties of groups rather than an intrinsic notion of random groups: (say in the density model) for each length ℓ we indeed define a measure μ_ℓ on the set of group presentations, but this measure does not converge as $\ell \to \infty$. Rather, for a given group property P, its probability of occurrence under μ_ℓ converges. As discussed in § I.4., the space of marked groups does not solve this problem because the notion of topological genericity in it is uninteresting (there are very different-looking connected components) and so an input from probability theory is required to know where in this space to look.

The *temperature model*, or *every-length density model* (discussed at the end of [Gro00], but already present in [Gro93], 9.B.(d)) attempts to solve these problems by directly defining a probability measure on the set of all (finite or infinite) group presentations, thus providing a well-defined notion of a random group. Note that this measure will project on the space of marked groups, and thus give access to the realm of infinitely presented groups.

As usual, fix a set of m generators, and consider the set F_m of reduced words. The principle is to construct a set of relators R by deciding at random, for each $r \in F_m$, whether we put it in R or not. Since there are much more long than short words, the probability to take r should decrease with the length of r. A very natural choice is to set

$$p(r) = (2m-1)^{(d-1)|r|}$$

where $|r|$ is the length of the word r, and $d \leqslant 1$ is a density parameter. Now, for each $r \in F_m$, with probability $p(r)$ we decide to put r as a relator in R (independently of what is decided for other r's). The random group is given by the presentation $G = \langle\, a_1, \ldots, a_m \mid R \,\rangle$.

Note that a priori G is *infinitely* presented.

Let us interpret the parameter d. The expected number of words in R with a given length ℓ is $2m(2m-1)^{\ell-1}(2m-1)^{(d-1)\ell}$ because there are $2m(2m-1)^{\ell-1}$ reduced words of length ℓ. Note that this behaves like $(2m-1)^{d\ell}$ (up to a benign constant $(2m)/(2m-1)$), which is of course very reminiscent of the density model.

In other words, if for each $\ell \in \mathbb{N}^*$, R'_ℓ is a random set of relators at density d and at length ℓ, then the union $R' = \bigcup_{\ell \in \mathbb{N}^*} R'_\ell$ has essentially the same probability law as R (up to replacement of an average number by a fixed number of relators, which for number such as $(2m-1)^\ell$ is negligible by the law of large numbers).

This justifies the name “all-length density model”. The “temperature” [Gro00] refers to the idea that a word $w \in F_m$ has “energy” $|w|$, and so if temperature is T the probability for a “random word” to be in “state” r (compared to its probability to be in state $r = e$) is $e^{-|r|/T}$, so that $T = 1/((1-d)\log(2m-1))$. The higher the temperature, the larger the set of relators R, the smaller the group G. When $T \to 0$, on the contrary, the set R “freezes” to the empty set so that $G = F_m$. Note that negative densities are meaningful in this model.

As an immediate consequence of the interpretation of d as a density, we get that if $d > 1/2$ (i.e. $T \geqslant 2/\log(2m-1)$) the group G is trivial with probability 1.

In this model, for each $d > -\infty$ there is a small but definite positive probability to pick all the generators $a_1, \ldots, a_m$ and put them as relators in R, in which case the group is trivial. So here we do not expect a phase transition between infinity and triviality of G with probabilities 0 and 1, but rather, a phase transition between a positive probability to be infinite and a zero probability to be infinite.

Up to this remark and the fact that the presentation is infinite at $d > 0$, the conjecture [Gro00] is the exact analogue of Theorem 11:

Conjecture.
If $d < 1/2$, the group G has a positive probability to be infinite, and more precisely to be a direct limit of infinite hyperbolic, torsion-free groups of geometric dimension 2.

Another way to express $d < 1/2$ is that the function $p(r)$ is in $\ell^2(F_m)$.

Not everything can happen with positive probability: for example at $d > 0$ we put an exponential number of generators, so that by a simple argument, with probability 1 the abelianization of G is trivial, and so Abelian groups never appear in this model (the support of the measure is not the whole space $\mathcal{G}_m$).

At $d < 0$, by the Borel-Cantelli lemma, the presentation for G is finite with probability 1, and the model is more or less related to the few-relator model with various lengths, so that for negative densities the conjecture follows from Theorem 5, and the group is even hyperbolic. But at $d \geqslant 0$ the presentation is infinite, and if the conjecture indeed holds the group will admit no finite presentation.

As a sidepoint, Theorem 27 implies property (T) for $d > 1/3$, with probability 1. This property is likely to happen much earlier.

An easy but important feature [Gro00] of this model is that for any $d \geqslant 0$, with probability 1 the group G has no finite quotient (compare the discussion above in § IV.e.). Indeed, let $\pi : G \to H$ be a finite quotient of G. The cosets $\pi^{-1}(h \in H)$ meet one element of F_m out of $\#H$, and so for $d \geqslant 0$ it is easy to see that R will contain one (actually infinitely many) element of each coset with probability 1, thus proving that $H = \{e\}$. This was for one single finite group H, but the union of countably many events of probability zero has probability zero again.

The main difficulty when dealing with the temperature model is failure of the local-global principle (see one possible statement in § V., Theorem 60, and other references a few sentences below), a.k.a. the Gromov-Cartan-Hadamard theorem, which allows to show hyperbolicity of a group by testing only isoperimetry for van Kampen diagrams of bounded size. This implies in particular that there exists an algorithm which, given a finite group presentation, answers positively when the group is hyperbolic (but may not stop if the group is not).

When the lengths in a group presentation are of very different orders of magnitude, this principle fails (or at least no suitable version of it is known). For a fixed density d, for any $\ell \in \mathbb{N}$ let $R_{\ell,d}$ be a random set of relators at density d and at length ℓ. Using the axioms in [Oll04] one can show that, for any constant $A \geqslant 1$, the group presented by

$$\left\langle a_1, \ldots, a_m \mid \bigcup_{\ell_0 \leqslant \ell \leqslant A\ell_0} R_{\ell,d} \right\rangle$$

is very probably hyperbolic, for large enough ℓ_0 depending on A. Then, using the theory of random quotients and iterating like in Proposition 43, for any $A \geqslant 1$ we can show that if $\ell_{i+1} \gg A\ell_i$, the group presented by

$$\left\langle a_1, \ldots, a_m \mid \bigcup_{i \in \mathbb{N}} \bigcup_{\ell_i \leqslant \ell \leqslant A\ell_i} R_{\ell,d} \right\rangle$$

will very probably be infinite and a direct limit of hyperbolic groups. But the techniques used to treat $\bigcup_{\ell_0 \leqslant \ell \leqslant A\ell_0} R_{\ell,d}$ are very different from those used to treat the passage from ℓ_i to ℓ_{i+1}, so that this "lacunarity" is currently needed (see [Gro03]). Note however that this lacunarity does not (at least explicitly) appear in Ol'shanskiĭ's treatment [Ols92] of the few-relator model with various lengths (thanks to the use of a homogeneous way to write isoperimetry as discussed in § IV.h. above).

So probably the key to the temperature model is a much better understanding of the local-global principle when the relators have very different lengths. The formulation given below (Theorem 60 in § V.) allows some looseness for the ratio of the lengths, and a careful exploitation of it results in replacing $\ell_0 \leqslant \ell \leqslant A\ell_0$ with $\ell_0 \leqslant \ell \leqslant \ell_0^\alpha$ in the above, for some exponent $\alpha > 1$. A first step would be to remove the dependency in ℓ_2/ℓ_1 in Theorem 60 [Oll-f]. But this is not enough to tackle the temperature model. The geometrizing of Cayley complexes as discussed in § IV.i. (after [Gro03], [DG] etc.) will also certainly be a key ingredient.

We refer the reader to [Gro87] (2.3.F and 6.8.M) and to [Bow91, Ols91b, Bow95, Pap96, DG, Oll-f] for more information on this important topic. There is not even a single statement unifying the various versions of the local-global principle written so far...

The same game can be played replacing F_m with any (especially hyperbolic!) initial group G_0 and killing random elements of G_0 according to the temperature scheme, thus transposing in this model all the random quotient questions of § IV.f., and endowing some neighborhood of each group in $\mathcal{G}_m$ with a canonical probability measure depending on density.

IV.l. Random Lie algebras. Ask Étienne Ghys about this (see also [Gro93], 9.B.(h)).

IV.m. Random Abelian groups, computer science and statistical physics. Phase transitions arose first in statistical physics and it is natural to ask whether the phase transition of random groups does model some physical phenomenon. The answer is presently unknown.

A fundamental problem of computer science is the 3-SAT problem, which asks whether a given set of clauses on Boolean variables can be satisfied. Each clause is of the form $(\neg)x_i \,\mathrm{OR}\, (\neg)x_j \,\mathrm{OR}\, (\neg)x_k$, where $(\neg)$ denotes optional negations and where $1 \leqslant i, j, k \leqslant n$. A set of clauses is *satisfiable* if each variable can be assigned the value true or false such that all clauses become true Boolean formulae. Variants exist in which the length of the clauses is not necessarily equal to 3. This problem is very important, and in particular it is NP-complete.

A widely used approach consists in observing the behavior of this problem for *random* choices of the clauses, for which methods from statistical physics are very useful (see e.g. [BCM02, MMZ01] for an introduction). In this context there is a phase transition depending on the ratio of the number of clauses to the number of Boolean variables: when this ratio is below a precise threshold the set of clauses is very probably satisfiable, whereas it is not above the threshold. Moreover, away from the threshold, naive algorithms perform very well though the problem is NP-complete.

This immediately brings to mind the triangular model of random groups (§ I.3.g.), which consists in taking relations of the form $x_i^{\pm 1} x_j^{\pm 1} x_k^{\pm 1} = e$ at random and asking whether the group presented by the elements $x_1, \ldots, x_n$ subject to these relations is trivial or not. This triangular model looks strikingly like a kind of "non-commutative" version of 3-SAT.

A commonly studied toy version of the 3-SAT problem is the XOR-SAT problem, using exclusive OR's instead of OR's in the clauses. This one has a polynomial-time solution (it reduces to a linear system modulo 2), hence is considerably simpler theoretically, but nevertheless seems to keep lots of interesting properties of 3-SAT. It can be interpreted as a random quotient of the commutative group $(\mathbb{Z}/2\mathbb{Z})^n$ (or sparse random matrices), thus in line with the intuition that the triangular model is a somewhat non-commutative random 3-SAT problem.

Random 3-SAT also exhibits phases: of course the satisfiability vs. non-satisfiability phases parallel the hyperbolicity vs. triviality phases for group, but moreover, the satisfiability phase breaks into two quite differently-behaved subphases, one in which the set of admissible truth value assignments to the variables is strongly connected and satisfiability is easy, and one in which the set of admissible truth value assignments breaks into many well-separated clusters. These two subphases evoke the freeness vs. (T) transition in the triangular model (Proposition 30 and Theorem 31): below this frontier, the group is infinite for trivial reasons, whereas above it, it is still infinite but not trivially so (compare performance of the group algorithms discussed in I.3.h.). This suggests that isolation of clusters of SAT solutions parallels isolation of the trivial representation among unitary representations of the group (one possible definition of property (T)).

The many possible assignments of truth values to the variables suggest to look not only at *the* random group given by a random presentation, but at all groups generated by the same elements and satisfying the random relations in the presentation (which are exactly the quotients of this group). Maybe the connectedness vs. many-clustering of solutions of SAT translates into some geometric property of the set of those groups, considered in the space $\mathcal{G}_m$ of marked groups (§ I.4., § IV.g.).

This is quite speculative and there may also be no relation at all between these fields. Nevertheless, methods from statistical physics and random-oriented computer science are certainly interesting tools to study for random group theorists.

Part V.
Proof of the density one half theorem

V.a. Prolegomena. Recall from the *Primer to geometric group theory* that, given a group presentation, a van Kampen diagram is basically a connected planar graph each oriented edge of which bears a generator of the presentation or its inverse (with opposite edges bearing inverse generators), such that the word labelling the boundary path of each face is (a cyclic permutation of) a relator in the presentation or its inverse. The diagram is said to be *reduced* if moreover some kind of trivial construction is avoided. We refer to [LS77, Ols91a, Rot95] for precise definitions. The set of reduced words that are read on the external boundary path of some van Kampen diagram coincides with the set of reduced words representing the trivial element in the group.

It is known [Gro87, Sho91a] that a group G is hyperbolic if and only if there exists a constant C such that any reduced word w representing the trivial element of G appears on the boundary of some van Kampen diagram with at most $C\,|w|$ faces.

In particular, to establish hyperbolicity it is enough to prove that there exists a constant $\alpha > 0$ such that for any diagram D, we have $|\partial D| \geqslant \alpha\,|D|$ (where $|D|$ is the number of faces of D and $|\partial D|$ the length of the boundary path of D [4]). This implies the above with $C = 1/\alpha$. Note that since reducing a van Kampen diagram preserves the boundary word, it is enough to check $|\partial D| \geqslant \alpha\,|D|$ for reduced diagrams (this would actually never hold for all non-reduced diagrams).

We are going to show that for a random group at density d and at length ℓ, with overwhelming probability any reduced van Kampen diagram satisfies $|\partial D| \geqslant (1-2d-\varepsilon)\ell\,|D|$ (i.e. we actually prove Theorem 13).

The idea is very nicely explained in [Gro93], 9.B. Remember the discussion of Gromov's density (§ I.2.): The probability that two random reduced words share a common initial subword[5] of length L is $1/(2m-1)^L$. So at density d, the probability that, in a set R made of $(2m-1)^{d\ell}$ random relators, *there exist* two words sharing a common initial subword of length L, is at most $(2m-1)^{2d\ell}(2m-1)^{-L}$ (this was Proposition 10).

The geometric way to think about it is to visualize a 2-face van Kampen diagram in which two faces of boundary length ℓ share L common edges. We have shown that the probability that two relators in R make such a diagram is at most $(2m-1)^{2d\ell-L}$ (up to an unimportant, subexponential factor $4\ell^2$ accounting for the positioning and orientation of the relators in the diagram).

Now consider a random presentation $\langle\, a_1, \ldots, a_m \mid R \,\rangle$ where R is made of $(2m-1)^{d\ell}$ random reduced words of length ℓ. Let D be any van Kampen diagram made by the relators in R. Each internal edge of D (i.e. an edge adjacent to two faces) forces

[4] which is not exactly the number of edges of ∂D in case the interior of D is not connected.

[5] Here and throughout the following we neglect the fact that for the first letter of a reduced word, we have $2m$ choices instead of $2m-1$ as for all subsequent letters; when dealing with cyclically reduced words, we also neglect the fact that for the last letter there may be $2m-1$ or $2m-2$ choices.

an equality between two letters of the two relators read on the two adjacent faces; for random reduced words this equality has a probability $1/(2m-1)$ to be fulfilled. So if L is the total number of internal edges in D, the probability that $|D|$ random reduced words fulfill the L constraints imposed by D is at most $1/(2m-1)^L$ (if the constraints are independent). So the probability that we can find $|D|$ relators in R fulfilling the constraints of D is at most $(2m-1)^{|D|d\ell}(2m-1)^{-L}$.

Choose any $\varepsilon > 0$. If $L > (d+\varepsilon)\,|D|\,\ell$, then the probability that D appears as a van Kampen diagram of the presentation R is less than $(2m-1)^{-\varepsilon|D|\ell}$ by the reasoning above, and so when $\ell \to \infty$, with overwhelming probability D does not appear as a van Kampen diagram of the random group. So we can assume $L \leqslant (d+\varepsilon)\,|D|\,\ell$.

Now we have

$$|\partial D| \geqslant |D|\,\ell - 2L$$

since D has $|D|$ faces, each of length ℓ, and each gluing between two faces decreases the boundary length by 2. (Equality occurs when the interior of D is connected; otherwise, "filaments" linking clusters of faces still increase boundary length.) Consequently, using $L \leqslant (d+\varepsilon)\,|D|\,\ell$ we get

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

as needed.

There are several obscure points in this proof. First, we did not justify why the constraints imposed by a van Kampen diagrams on letters of the presentation can be supposed to be independent (in fact, they are not as soon as the diagram involves several times the same relator[6]), so we are a priori not allowed to multiply all probabilities involved as we did. Second, we should exclude simultaneously *all* diagrams violating the isoperimetric inequality, and we only estimated the probability that one particular diagram is excluded. Third, note that the trivial group, as well as any finite group, is hyperbolic and thus satisfies the isoperimetric inequality, so we have proven that the group is hyperbolic but not necessarily infinite.

The latter is treated by a cohomological dimension argument, see below. The second problem is solved using the local-global principle of hyperbolic geometry (or Gromov-Cartan-Hadamard theorem) which will be explained later. The first point requires a more in-depth study of the probability for random relators to fulfill a diagram, which we now turn to.

[6] The proof given in [Żuk03] for the triangular model is partly incorrect too, but in a more subtle way when a diagram involves several copies of a relator glued to itself. Namely, on page 659 of [Żuk03]: "First put in the diagram n_1 relators r_1. If they have some edges in common, denote by l_1 the length of the longest common sequence, i.e. $0 \leqslant l_1 \leqslant 3$" and then it is stated that, given the constraints of the diagram, the number of choices for such a relator is at most $(2m-1)^{3-l_1}$.

Either l_1 denotes the maximal length of the intersection of two faces bearing r_1. In this case it is not true that the total number L of internal edges of the diagram is at most $\sum n_i l_i$.

Or l_1 denotes the maximal length of the intersection of a face with the union of all other faces bearing r_1. Then let D be a 3-face diagram bearing three copies of r_1, the second copy having reverse orientation, and with the second letter of the first copy glued to the first letter of the second copy, and the second letter of the second copy glued to the first letter of the third copy. In this case $l_1 = 2$ but the number of choices for r_1 is $(2m-1)^2$.

V.b. Probability to fulfill a diagram. First, we need a precise definition of what it means for random words to fulfill a van Kampen diagram. We define an abstract diagram to be a van Kampen diagram in which we forget the actual relators associated with the faces, but only remember the geometry of the diagram, which faces bear the same relator as each other, the orientation of the relator of each face, and where the relators begin. Namely:

Definition 57.
An abstract diagram $\mathcal{D}$ *is a connected planar graph without valency-*1 *vertices, equipped with the following data:*

- *An integer* $1 \leqslant n \leqslant |\mathcal{D}|$ *called the* number of distinct relators in $\mathcal{D}$ *(where* $|\mathcal{D}|$ *is the number of faces of* $\mathcal{D}$*);*
- *A surjective map from the faces of* $\mathcal{D}$ *to the set* $\{1, 2, \ldots, n\}$*; a face with image* i *is said to* bear relator i*;*
- *For each face* f*, a distinguished edge on the boundary of* f *and an orientation of the plane* ± 1*; if* p *is the boundary path of* f *with the distinguished edge as first edge and oriented according to the orientation of* f*, we call the* k*-th edge of* p *the* k-th edge of f.

An n*-tuple* $(w_1, \ldots, w_n)$ *of cyclically*[7] *reduced words is said to* fulfill $\mathcal{D}$ *if the following holds: for each two faces* f_1 *and* f_2 *bearing relators* i_1 *and* i_2*, such that the* k_1*-th edge of* f_1 *is equal to the* k_2*-th edge of* f_2*, then the* k_1*-th letter of* w_{i_1} *and the* k_2*-th letter of* w_{i_2} *are inverse (when the orientations of* f_1 *and* f_2 *agree) or equal (when the orientations disagree). For a* n'*-tuple of words with* $n' \leqslant n$*, define a* partial fulfilling *similarly.*

An abstract diagram is said to be reduced *if no edge is adjacent to two faces bearing the same relator with opposite orientations such that the edge is the* k*-th edge of both faces.*

In other words, putting w_i on the faces of $\mathcal{D}$ bearing relator i turns $\mathcal{D}$ into a genuine van Kampen diagram.

It is clear that conversely, any van Kampen diagram defines an associated abstract diagram (which is unique up to reordering the relators). A van Kampen diagram is reduced if and only if its associated abstract diagram is.

We can choose the order of enumeration of the relators and in particular we can ask that the number of faces bearing relator i is non-increasing with i (call relator 1 the most frequent relator, etc.).

Note that a face of a graph can be non-trivially adjacent to itself, in which case we have $f_1 = f_2$ above (but then of course $k_1 \neq k_2$).

[7] Here we work with cyclically reduced words to avoid the following technical annoyance: the beginning and end of a reduced word may cancel, which forces to consider van Kampen diagrams with "inward spurs" in some faces. Anyway the theorem holds for any version, since the "probabilistic cost" of such a cancellation is identical to the probabilistic cost of a cancellation between two relators. Note however that the same theorem does not hold for plain (non-reduced) random words, since then the combinatorics of possible cancellations is exponential, and the critical density is lower than 1/2, as explained in § II.2.a. and [Oll04].

Hereafter we limit ourselves to abstract diagrams each face of which has boundary path of length ℓ, in accordance with the density model of random groups. Our goal is to prove the following:

Proposition 58.
Let R be a random set of relators at density d and at length ℓ. Let D be a reduced abstract diagram and let $\varepsilon > 0$.

Then either $|\partial D| \geqslant |D|\,\ell(1-2d-2\varepsilon)$, or the probability that there exists a tuple of relators in R fulfilling D is less than $(2m-1)^{-\varepsilon\ell}$.

Note that in the "intuitive" proof above, we had a probability $(2m-1)^{-\varepsilon|D|\ell}$ instead.

To prove this proposition we shall need some more definitions. Let n be the number of distinct relators in D. For $1 \leqslant i \leqslant n$ let m_i be the number of times relator i appears in D. As mentioned above, up to reordering the relators we can suppose that $m_1 \geqslant m_2 \geqslant \ldots \geqslant m_n$.

For $1 \leqslant i_1, i_2 \leqslant n$ and $1 \leqslant k_1, k_2 \leqslant \ell$ say that $(i_1, k_1) > (i_2, k_2)$ if $i_1 > i_2$, or if $i_1 = i_2$ and $k_1 > k_2$. Let e be an edge of D adjacent to faces f_1 and f_2 bearing relators i_1 and i_2, which is the k_1-th edge of f_1 and the k_2-th edge of f_2. Say edge e *belongs to* f_1 if $(i_1, k_1) > (i_2, k_2)$, and *belongs to* f_2 if $(i_2, k_2) > (i_1, k_1)$, so that an edge belongs to the second face it meets.

Note that since D is reduced, each internal edge belongs to some face: indeed if $(i_1, k_1) = (i_2, k_2)$ then either the two faces have opposite orientations and then D is not reduced (by definition), or they have the same orientation and the diagram is never fulfillable since a letter would have to be its own inverse.

Let $\delta(f)$ be the number of edges belonging to face f. For $1 \leqslant i \leqslant n$ let

$$\delta_i = \max_{f \text{ face bearing } i} \delta(f) \tag{1}$$

which will intuitively measure the "log-probabilistic cost" of relator i (lemma below).

Since each internal edge belongs to some face, we have

$$|\partial D| \geqslant \ell\,|D| - 2 \sum_{f \text{ face of } D} \delta(f) \geqslant \ell\,|D| - 2 \sum_{1 \leqslant i \leqslant n} m_i \delta_i \tag{2}$$

Lemma 59.
For $1 \leqslant i \leqslant n$ let p_i be the probability that i randomly chosen cyclically reduced words $w_1, \ldots, w_i$ partially fulfill D (and $p_0 = 1$). Then

$$p_i / p_{i-1} \leqslant (2m-1)^{-\delta_i} \tag{3}$$

The lemma is proven as follows: Suppose that $i-1$ words $w_1, \ldots, w_{i-1}$ partially fulfilling D are given. Then, successively choose the letters of the word w_i in a way to fulfill the diagram. Let f be a face of D bearing relator i and realizing the maximum δ_i.

Let $k \leqslant \ell$ and suppose the first $k-1$ letters of w_i are chosen. If the k-th edge of f belongs to f, then this means that the other face f' meeting this edge either bears a relator $i' < i$, or bears i too but the edge appears as the $k' < k$-th edge in f' (it may even happen that $f' = f$). In both cases, in order to fulfill the diagram the k-th letter of w_i is imposed by the letter already present on the edge, so that choosing it at random has a probability $1/(2m-1)$ to be correct[8]. The lemma is proven.

Now for $1 \leqslant i \leqslant n$ let P_i be the probability that there exists a i-tuple of words partially fulfilling D in the random set of relators R. We trivially[9] have

$$P_i \leqslant (\#R)^i p_i = (2m-1)^{id\ell} p_i \tag{4}$$

and according to the density philosophy, $id\ell + \log_{2m-1} p_i$ is to be seen as the dimension of the i-tuples of relators partially fulfilling D (i.e. the log of the expected number of such i-tuples). This explains the role played by logs in the few next lines—beware these logs are negative!

Combining Equations (2) and (3) we get

$$\begin{aligned} |\partial D| &\geqslant \ell\,|D| + 2\sum m_i \left(\log_{2m-1} p_i - \log_{2m-1} p_{i-1}\right) && (5)\\ &= \ell\,|D| + 2\sum (m_i - m_{i+1}) \log_{2m-1} p_i && (6) \end{aligned}$$

and Equation (4) yields (here we use $m_i \geqslant m_{i+1}$)

$$|\partial D| \geqslant \ell\,|D| + 2\sum (m_i - m_{i+1})(\log_{2m-1} P_i - id\ell) \tag{7}$$

and observe here that $\sum (m_i - m_{i+1}) id\ell = d\ell \sum m_i = d\ell\,|D|$, hence

$$|\partial D| \geqslant \ell\,|D|\,(1-2d) + 2\sum (m_i - m_{i+1}) \log_{2m-1} P_i \tag{8}$$

so that setting $P = \min_i P_i$ (and using $m_i \geqslant m_{i+1}$ again) we get

$$\begin{aligned} |\partial D| &\geqslant \ell\,|D|\,(1-2d) + 2(\log_{2m-1} P) \sum (m_i - m_{i+1}) && (9)\\ &= \ell\,|D|\,(1-2d) + 2m_1 \log_{2m-1} P && (10)\\ &\geqslant |D| \left(\ell(1-2d) + 2\log_{2m-1} P\right) && (11) \end{aligned}$$

since $m_1 \leqslant |D|$.

Of course a diagram is fulfillable if and only if it is partially fulfillable for any $i \leqslant n$ and so

$$\Pr\left(D \text{ is fulfillable by relators of } R\right) \leqslant P \leqslant (2m-1)^{\frac{1}{2}(|\partial D|/|D| - \ell(1-2d))} \tag{12}$$

which was to be proven.

[8] See footnote 7.

[9] Here it is even true that $P_i/P_{i-1} \leqslant (\#R) p_i/p_{i-1}$, because p_i/p_{i-1} is independent of the value of the words $w_1, \ldots, w_{i-1}$. But this is no longer true in more general contexts such as random quotients of hyperbolic groups, where one has to condition by some properties of $w_1, \ldots, w_{i-1}$ (the "apparent lengths" in [Oll04]).

V.c. The local-global principle, or Gromov-Cartan-Hadamard theorem. The proof above applies only to one van Kampen diagram. But a deep result of Gromov ([Gro87], 2.3.F, 6.8.M) states that hyperbolicity of a space can be tested on balls of finite radius. This somehow generalizes the classical Cartan-Hadamard theorem stating that a simply connected complete Riemannian manifold with non-positive sectional curvature is homeomorphic to $\mathbb{R}^n$.

This implies in particular that hyperbolicity is semi-testable in the sense that there exists an algorithm which, given a presentation of a hyperbolic group, outputs an upper bound for the hyperbolicity constant (but which may not stop for non-hyperbolic presentations). Such an algorithm has indeed been implemented [EH01, Hol95].

Following Gromov, the principle has been given various, effective or non-effective formulations [Bow91, Ols91b, Bow95, Pap96, DG, Oll-f]. The variant best suited to our context is the following [Oll-f]:

Theorem 60.
Let $G = \langle a_1, \ldots, a_m \mid R \rangle$ be a finite group presentation and let ℓ_1, ℓ_2 be the minimal and maximal lengths of a relator in R.

For a van Kampen diagram D with respect to the presentation set

$$\mathcal{A}(D) = \sum_{f \text{ face of } D} |\partial f|$$

where $|\partial f|$ is the length of the boundary path of face f.

Let $C > 0$. Choose $\varepsilon > 0$. Suppose that for some K greater than $10^{50}\,(\ell_2/\ell_1)^3\,\varepsilon^{-2}\,C^{-3}$, any reduced[10] van Kampen diagram D with $\mathcal{A}(D) \leqslant K\ell_2$ satisfies

$$|\partial D| \geqslant C\,\mathcal{A}(D)$$

Then any reduced van Kampen diagram D satisfies

$$|\partial D| \geqslant (C - \varepsilon)\,\mathcal{A}(D)$$

and in particular the group is hyperbolic.

Back to random groups. Here all relators in the presentation have the same length ℓ, so that $\mathcal{A}(D) = \ell\,|D|$. In particular, the assumption $\mathcal{A}(D) \leqslant K\ell_2$ in the theorem becomes $|D| \leqslant K$, i.e. we have to check diagrams with at most K faces.

Choose any $\varepsilon > 0$. Set $K = 10^{50}\,\varepsilon^{-2}\,(1-2d-2\varepsilon)^{-3}$, which most importantly does not depend on ℓ. Let $N(K,\ell)$ be the number of abstract diagrams with K faces all of which have their boundary path of length ℓ. It can easily be checked (using the Euler formula) that for fixed K, $N(K,\ell)$ grows polynomially with ℓ (a rough estimate yields $N(K,\ell) \leqslant \ell^{4K} N(K)$).

We know (Proposition 58) that for any reduced abstract diagram D fixed in advance and violating the inequality $|\partial D| \geqslant (1 - 2d - 2\varepsilon)\ell\,|D|$, the probability that it appears as a van Kampen diagram of the presentation is at most $(2m-1)^{-\varepsilon\ell}$. So the

[10] This constraint can be weakened.

probability that there exists a reduced van Kampen diagram with at most K faces, violating the inequality $|\partial D| \geqslant (1-2d-2\varepsilon)\ell\,|D|$, is less than $N(K,\ell)(2m-1)^{-\varepsilon\ell}$. But, for fixed K and ε, this tends to 0 when $\ell \to \infty$ since $N(K,\ell)$ grows subexponentially with ℓ.

So with overwhelming probability, all reduced diagrams of the presentation with at most K faces satisfy the isoperimetric inequality $|\partial D| \geqslant (1-2d-2\varepsilon)\ell\,|D|$. Applying Theorem 60 (with our choice of K) yields that all reduced van Kampen diagrams D satisfy $|\partial D| \geqslant (1-2d-3\varepsilon)\ell\,|D|$ as needed.

The size of the constant K and the large value of $N(K,\ell)$ may explain why computer experiments (§ I.3.h.) found the group to be trivial too often...

V.d. Infiniteness. The isoperimetric inequality above is shown to hold for any reduced van Kampen diagram (and not only for one van Kampen diagram per boundary word, which is enough to be hyperbolic). This implies in particular that there is no spherical diagram (a spherical diagram being a limit case of planar diagram of zero boundary length, thus violating the isoperimetric inequality) and so the Cayley 2-complex is aspherical[11], hence the group has geometric (hence cohomological) dimension at most 2. Any group with torsion has infinite cohomological dimension, and so the random group is torsion-free (which rules out non-trivial finite groups).

The trivial group is excluded since, using asphericity of the Cayley complex, the Euler characteristic of the group is equal to $1-m+\#R = 1-m+(2m-1)^{d\ell}$; for positive d this is >1, whereas the trivial group has Euler characteristic 1 (and excluding the trivial group for $d>0$ excludes it a fortiori for $d=0$). The elementary hyperbolic group $\mathbb{Z}$ is excluded for the same reason.

[11] For this to work one needs a careful definition of van Kampen diagrams, since there are several non-equivalent notions of asphericity. See e.g. the discussion in [Oll-a].

References

[And58] R. D. Anderson, *A characterization of the universal curve and a proof of its homogeneity*, Ann. Math. **67** (1958), n°2, 313–324.

[Ar97] G. N. Arzhantseva, *On groups in which subgroups with a fixed number of generators are free* (Russian), Fundam. Prikl. Mat. **3** (1997), n°3, 675–683.

[Ar98] G. N. Arzhantseva, *Generic properties of finitely presented groups and Howson's theorem*, Comm. Alg. **26** (1998), n°4, 3783–3792.

[Ar00] G. N. Arzhantseva, *A property of subgroups of infinite index in a free group*, Proc. Amer. Math. Soc. **128** (2000), n°11, 3205–3210.

[AC04] G. N. Arzhantseva, P.-A. Cherix, *On the Cayley graph of a generic finitely presented group*, Bull. Belg. Math. Soc. **11** (2004), n°4, 589–601.

[AL02] G. N. Arzhantseva, I. G. Lysenok, *Growth tightness for word hyperbolic groups*, Math. Z. **214** (2002), n°3, 597–611.

[AO96] G. N. Arzhantseva, A. Yu. Ol'shanskiĭ, *The class of groups all of whose subgroups with lesser number of generators are free is generic*, Mat. Zametki **59** (1996), n°4, 489–496; English translation in Math. Notes **59** (1996), n°3–4, 350–355.

[BCM02] G. Biroli, S. Cocco, R. Monasson, *Phase transitions and complexity in computer science: an overview of the statistical physics approach to the random satisfiability problem*, Physica A **306** (2002), 381–394.

[BH99] M. R. Bridson, A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der mathematischen Wissenschaften **319**, Springer (1999).

[BHV] B. Bekka, P. de la Harpe, A. Valette, *Kazhdan groups*, book in preparation.

[BM91] M. Bestvina, G. Mess, *The boundary of negatively curved groups*, J. Amer. Math. Soc. **4** (1991), n°3, 469–481.

[Bow91] B. H. Bowditch, *Notes on Gromov's hyperbolicity criterion for path-metric spaces*, in *Group theory from a geometrical viewpoint*, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991), 64–167.

[Bow95] B. H. Bowditch, *A short proof that a subquadratic isoperimetric inequality implies a linear one*, Michigan Math. J. **42** (1995), n°1, 103–107.

[BŚ97] W. Ballmann, J. Świątkowski, *On L^2-cohomology and property (T) for automorphism groups of polyhedral cell complexes*, GAFA, Geom. Funct. Anal. **7** (1997), 615–645.

[BSz] I. Belegradek, A. Szczepański, *Endomorphisms of relatively hyperbolic groups*, preprint, ArXiv `math.GR/0501321`

[Bum04] I. Bumagina, *The Hopf property for subgroups of hyperbolic groups*, Geom. Dedicata **106** (2004), 211–230.

[BW05] I. Bumagin, D. T. Wise, *Every group is an outer automorphism group of a finitely generated group*, J. Pure Appl. Algebra **200** (2005), n°1–2, 137–147.

[CCJJV01] P.-A. Cherix, M. Cowling, P. Jolissaint, P. Julg, A. Valette, *Groups with the Haagerup property. Gromov's a-T-menability*, Progress in Math. **197**, Birkhäuser (2001).

[CDP90] M. Coornaert, T. Delzant, A. Papadopoulos, *Les groupes hyperboliques de Gromov*, Lecture Notes in Mathematics **1441**, Springer (1990).

[CG05] Ch. Champetier, V. Guirardel, *Limit groups as limits of free groups*, Israel J. Math. **146** (2005), 1–75.

[Ch91] Ch. Champetier, *Propriétés génériques des groupes de présentation finie*, Ph.D. Thesis, Université Lyon 1 (1991).

[Ch93] Ch. Champetier, *Cocroissance des groupes à petite simplification*, Bull. London Math. Soc. **25** (1993), n°5, 438–444.

[Ch94] Ch. Champetier, *Petite simplification dans les groupes hyperboliques*, Ann. Fac. Sci. Toulouse Math., ser. 6, vol. 3 (1994), n°2, 161–221.

[Ch95] Ch. Champetier, *Propriétés statistiques des groupes de présentation finie*, J. Adv. Math. **116** (1995), n°2, 197–262.

[Ch00] Ch. Champetier, *L'espace des groupes de type fini*, Topology **39** (2000), n°4, 657–680.

[CM88] A. Connes, H. Moscovici, *Conjecture de Novikov et groupes hyperboliques*, C. R. Acad. Sci. Paris Sér. I Math. **307** (1988), n°9, 475–480.

[CM90] A. Connes, H. Moscovici, *Cyclic cohomology, the Novikov conjecture and hyperbolic groups*, Topology **29** (1990), n°3, 345–388.

[CMV04] P.-A. Cherix, F. Martin, A. Valette, *Spaces with measured walls, the Haagerup property and property* (T), Ergodic Theory Dynam. Systems **24** (2004), n°6, 1895–1908.

[CN] I. L. Chatterji, G. A. Niblo, *From wall spaces to* $CAT(0)$ *cube complexes*, to appear in Internat. J. Algebra Comput. special issue, ArXiv `math.GT/0309036`

[Coh82] J. M. Cohen, *Cogrowth and amenability of discrete groups*, J. Funct. Anal. **48** (1982), 301–309.

[Cor-a] Y. de Cornulier, *Finitely presentable, non-Hopfian groups with Kazhdan's Property* (T) *and infinite outer automorphism group*, to appear in Proc. Amer. Math. Soc., ArXiv `math.GR/0502140`

[Cor-b] Y. de Cornulier, *A note on quotients of word hyperbolic groups with Property* (T), preprint, ArXiv `math.GR/0504193`

[CS98] P.-A. Cherix, G. Schaeffer, *An asymptotic Freiheitssatz for finitely generated groups*, Enseign. Math. (2) **44** (1998), n°1–2, 9–22.

[CV96] P.-A. Cherix, A. Valette, *On spectra of simple random walks on one-relator groups*, Pacific J. Math. **175** (1996), n°2, 417–438.

[Del96a] T. Delzant, *Sous-groupes distingués et quotients des groupes hyperboliques*, Duke Math. J. **83** (1996), n°3, 661–682.

[Del96b] T. Delzant, *Décomposition d'un groupe en produit libre ou somme amalgamée*, J. Reine Angew. Math. **470** (1996), 153–180.

[Del-a] T. Delzant, *Asphéricité et sous-groupes libres des groupes hyperboliques*, manuscript (1996).

[Del-b] T. Delzant, *Mesoscopic curvature and very small cancellation theory*, manuscript (2002).

[DG] T. Delzant, M. Gromov, *Groupes de Burnside et géométrie hyperbolique*, preprint.

[Dix02] J. D. Dixon, *Probabilistic group theory*, C. R. Math. Acad. Sci. Soc. R. Can. **24** (2002), n°1, 1–15.

[DSV03] G. Davidoff, P. Sarnak, A. Valette, *Elementary number theory, group theory, and Ramanujan graphs*, London Mathematical Society Student Texts **55**, Cambridge University Press, Cambridge (2003).

[EH01] D. B. A. Epstein, D. Holt, *Computation in word-hyperbolic groups*, Internat. J. Algebra Comput. **11** (2001), n°4, 467–487.

[EH] D. B. A. Epstein, D. Holt, *The linearity of the conjugacy problem in word hyperbolic groups*, to appear in Internat. J. Algebra Comput.

[GhH90] É. Ghys, P. de la Harpe, *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math. **83**, Birkhäuser (1990).

[Ghy90] É. Ghys, *Les groupes hyperboliques*, séminaire Bourbaki **722** (1990). Published in Astérisque **189–190** (1990), 203–238.

[Ghy03] É. Ghys, *Groupes aléatoires*, séminaire Bourbaki **916** (2003). Published in Astérisque **294** (2004), 173–204.

[GH97] R. I. Grigorchuk, P. de la Harpe, *On problems related to growth, entropy, and spectrum in group theory*, Dynam. Control Systems **3** (1997), n°1, 51–89.

[Gri80] 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.

[Gri84] R. I. Grigorchuk, *Degrees of growth of finitely generated groups and the theory of invariant means*, Izv. Akad. Nauk SSSR Ser. Mat. **48** (1984), n°5, 939–985; English translation in Math. USSR—Izvestiya **25** (1985), n°2, 259–300.

[Gro78] M. Gromov, *Hyperbolic manifolds, groups and actions*, in *Riemann surfaces and related topics. Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978)*, Ann. of Math. Stud. **97**, Princeton University Press, Princeton (1981), 183–213.

[Gro81] M. Gromov, *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Publ. Math. **53** (1981), 53–73.

[Gro83] M. Gromov, *Infinite groups as geometric objects*, in *Proceedings of the International Congress of Mathematicians*, Vol. 1, 2, Warsaw (1983), 385–392, PWN, Warsaw, 1984.

[Gro87] M. Gromov, *Hyperbolic groups*, in *Essays in group theory*, ed. S. M. Gersten, Springer (1987), 75–265.

[Gro93] M. Gromov, *Asymptotic invariants of infinite groups*, in *Geometric group theory*, ed. G. Niblo, M. Roller, Cambridge University Press, Cambridge (1993).

[Gro00] M. Gromov, *Spaces and questions*, GAFA 2000 (Tel Aviv, 1999), Geom. Funct. Anal. Special Volume (2000), Part I, 118–161.

[Gro01a] M. Gromov, *Mesoscopic curvature and hyperbolicity*, in *Global differential geometry: the mathematical legacy of Alfred Gray*, ed. M. Fernández, J. A. Wolf, Contemporary Mathematics **288**, American Mathematical Society (2001), 58–69.

[Gro01b] M. Gromov, CAT(κ)*-spaces: construction and concentration*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov (POMI) **280** (2001), Geom. i Topol. **7**, 100–140, 299–300; English translation in J. Math. Sci. (N. Y.) **119** (2004), n°2, 178–200.

[Gro01c] M. Gromov, *Small cancellation, unfolded hyperbolicity, and transversal measures*, in *Essays on geometry and related topics, Mémoires dédiés à*

André Haefliger, ed. É. Ghys, P. de la Harpe, V. F. R. Jones, V. Sergiescu, T. Tsuboi, Monograph. Enseign. Math. **38**, L'Enseignement mathématique, Genève (2001), vol. 2, 371–399.

[Gro03] M. Gromov, *Random walk in random groups*, GAFA, Geom. Funct. Anal. **13** (2003), n°1, 73–146.

[Gup89] N. Gupta, *On groups in which every element has finite order*, Amer. Math. Monthly **96** (1989), n°4, 297–308.

[Guy] L. Guyot, *Estimations de dimensions de Minkowski dans l'espace des groupes marqués*, preprint.

[Har00] P. de la Harpe, *Topics in geometric group theory*, Chicago University Press (2000).

[Hig98] N. Higson, *The Baum-Connes conjecture*, in *Proceedings of the International Congress of Mathematicians*, Vol. II (Berlin, 1998), Doc. Math. 1998, Extra Vol. II, 637–646 (electronic).

[HLS02] N. Higson, V. Lafforgue, G. Skandalis, *Counterexamples to the Baum-Connes conjecture*, GAFA, Geom. Funct. Anal. **12** (2002), n°2, 330–354.

[Hol95] D. Holt, *KBMAG—Knuth-Bendix in Monoids and Automatic Groups*, software package (1995), available at `http://homepages.maths.warwick.ac.uk/~dfh/`

[Hol00] D. Holt, *Word-hyperbolic groups have real-time word problem*, Internat. J. Algebra Comput. **10** (2000), n°2, 221–227.

[HP98] F. Haglund, F. Paulin, *Simplicité de groupes d'automorphismes d'espaces à courbure négative*, Geom. Topol. Monograph **1** (1998), 181–248.

[HV89] P. de la Harpe, A. Valette, *La propriété* (T) *de Kazhdan pour les groupes localement compacts*, Astérisque **175**, Soc. Math. Fr. (1989).

[HW] G. C. Hruska, D. T. Wise, *Axioms for finiteness of cubulations*, preprint (2004).

[Iva94] S. V. Ivanov, *The free Burnside groups of sufficiently large exponents*, Internat. J. Algebra Comput. **4** (1994), n°1–2, 1–308.

[Iva98] S. V. Ivanov, *On the Burnside problem for groups of even exponent*, in *Proceedings of the International Congress of Mathematicians*, Vol. II (Berlin, 1998), Doc. Math. 1998, Extra Vol. II, 67–75 (electronic).

[IO96] S. V. Ivanov, A. Yu. Ol'shanskiĭ, *Hyperbolic groups and their quotients of bounded exponents*, Trans. Amer. Math. Soc. **348** (1996), n°6, 2091–2138.

[Kes59] H. Kesten, *Symmetric random walks on groups*, Trans. Amer. Math. Soc. **92** (1959), 336–354.

[KL05] M. Kreck, W. Lück, *The Novikov conjecture. Geometry and algebra*, Oberwolfach Seminars **33**, Birkhäuser (2005).

[KMSS03] I. Kapovich, A. Myasnikov, P. Schupp, V. Shpilrain, *Generic-case complexity, decision problems in group theory, and random walks*, J. Algebra **264** (2003), n°2, 665–694.

[KMSS05] I. Kapovich, A. Myasnikov, P. Schupp, V. Shpilrain, *Average-case complexity and decision problems in group theory*, Adv. Math. **190** (2005), n°2, 343–359.

[KS05] I. Kapovich, P. Schupp, *Genericity, the Arzhantseva-Ol'shanskiĭ method and the isomorphism problem for one-relator groups*, Math. Ann. **331** (2005), n°1, 1–19.

[KS] I. Kapovich, P. Schupp, *Delzant's T-invariant, Kolmogorov complexity and one-relator groups*, to appear in Comm. Math. Helvetici, ArXiv `math.GR/0305353`

[KSS] I. Kapovich, P. Schupp, V. Shpilrain, *Generic properties of Whitehead's algorithm and isomorphism rigidity of random one-relator groups*, to appear in Pacific J. Math., ArXiv `math.GR/0303386`

[Laf02] V. Lafforgue, *K-théorie bivariante pour les algèbres de Banach et conjecture de Baum-Connes*, Invent. Math. **149** (2002), n°1, 1–95.

[LS77] R. C. Lyndon, P. E. Schupp, *Combinatorial group theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete **89**, Springer (1977).

[Lub94] A. Lubotzky, *Discrete groups, expanding graphs and invariant measures*, Progress in Math. **125**, Birkhäuser (1994).

[MMZ01] O. C. Martin, R. Monasson, R. Zecchina, *Statistical mechanics methods and phase transitions in optimization problems. Phase transitions in combinatorial problems (Trieste, 1999)*, Theoret. Comput. Sci. **265** (2001), n°1–2, 3–67.

[MY02] I. Mineyev, G. Yu, *The Baum-Connes conjecture for hyperbolic groups*, Invent. Math. **149** (2002), n°1, 97–122.

[Nic04] B. Nica, *Cubulating spaces with walls*, Algebr. Geom. Topol. **4** (2004), 297–309 (electronic).

[NR98] G. A. Niblo, M. A. Roller, *Groups acting on cubes and Kazhdan's property (T)*, Proc. Amer. Math. Soc. **126** (1998), n°3, 693–699.

[Oll04] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, GAFA, Geom. Funct. Anal. **14** (2004), n°3, 595–679.

[Oll05a] Y. Ollivier, *Cogrowth and spectral gap of generic groups*, Ann. Inst. Fourier (Grenoble) **55** (2005), n°1, 289–317.

[Oll05b] Y. Ollivier, *Collapsing of random quotients of hyperbolic groups with torsion*, C.R. Math. Acad. Sci. Paris **341** (2005), 137–140.

[Oll-a] Y. Ollivier, *On a small cancellation theorem of Gromov*, to appear in Bull. Belg. Math. Soc., ArXiv `math.GR/0310022`

[Oll-b] Y. Ollivier, *Growth exponent of generic groups*, to appear in Comment. Math. Helv., ArXiv `math.GR/0401050`

[Oll-c] Y. Ollivier, *Cayley graphs containing expanders, after Gromov*, manuscript (2003).

[Oll-d] Y. Ollivier, *Spectral interpretations of Property* (T), manuscript (2003).

[Oll-e] Y. Ollivier, *Remarques sur la croissance et la cocroissance des groupes aléatoires*, informal notes (2004).

[Oll-f] Y. Ollivier, *Some small cancellation properties of random groups*, to appear in Internat. J. Algebra Comput., ArXiv `math.GR/0409226`

[Ols82] A. Yu. Ol'shanskiĭ, *The Novikov-Adyan theorem*, Mat. Sb. (N.S.) **118** (1982), n°2, 203–235. English translation in Math. USSR—Sb.

[Ols83] A. Yu. Ol'shanskiĭ, *On a geometric method in the combinatorial group theory*, in *Proceedings of the International Congress of Mathematicians*, Vol. 1, 2 (Warsaw, 1983), 415–424, PWN, Warsaw, 1984.

[Ols91a] A. Yu. Ol'shanskiĭ, *Geometry of defining relations in groups* (translated from the 1989 Russian original), Mathematics and its Applications (Soviet Series) **70**, Kluwer Academic (1991).

[Ols91b] A. Yu. Ol'shanskiĭ, *Hyperbolicity of groups with subquadratic isoperimetric inequality*, Internat. J. Algebra Comput. **1** (1991), n°3, 281–289.

[Ols91c] A. Yu. Ol'shanskiĭ, *Periodic quotient groups of hyperbolic groups*, Mat. Sb. **182** (1991), n°4, 543–567; English translation in Math. USSR—Sb. **72** (1992), n°2, 519–541.

[Ols92] A. Yu. Ol'shanskiĭ, *Almost every group is hyperbolic*, Internat. J. Algebra Comput. **2** (1992), n°1, 1–17.

[Ols93] A. Yu. Ol'shanskiĭ, *On residualing homomorphisms and G-subgroups of hyperbolic groups*, Internat. J. Algebra Comput. **3** (1993), n°4, 365–409.

[OW-a] Y. Ollivier, D. T. Wise, *Kazhdan groups with infinite outer automorphism group*, to appear in Trans. Amer. Math. Soc., ArXiv `math.GR/0409203`

[OW-b] Y. Ollivier, D. T. Wise, *Cubulating groups at density* 1/6, preprint.

[Pan98] P. Pansu, *Formules de Matsushima, de Garland et propriété* (T) *pour des groupes agissant sur des espaces symétriques ou des immeubles*, Bull. Soc. Math. Fr. **126** (1998), n°1, 107–139.

[Pan03] P. Pansu, *Groupes aléatoires*, lecture at the Soc. Math. Fr. annual meeting (2003), available at `http://www.math.u-psud.fr/~pansu/liste-prepub.html`

[Pap96] 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.

[Pau91] F. Paulin, *Outer automorphisms of hyperbolic groups and small actions on R-trees*, in *Arboreal group theory (Berkeley, CA, 1988)*, Math. Sci. Res. Inst. Publ. **19**, Springer, New York (1991), 331–343.

[Pau03] F. Paulin, *Sur la théorie élémentaire des groupes libres (d'après Sela)*, séminaire Bourbaki **922** (2003). Published in Astérisque **294** (2004), 363–402.

[Rip82] E. Rips, *Subgroups of small cancellation groups*, Bull. London Math. Soc. **14** (1982), n°1, 45–47.

[Rot95] J. J. Rotman, *An introduction to the theory of groups*, fourth edition, Graduate Texts in Mathematics **148**, Springer (1995).

[Sag95] M. Sageev, *Ends of group pairs and non-positively curved cube complexes*, Proc. London Math. Soc. (3) **71** (1995), n°3, 585–617.

[Sel92] Z. Sela, *Uniform embeddings of hyperbolic groups in Hilbert spaces*, Israel J. Math. **80** (1992), n°1–2, 171–181.

[Sel95] Z. Sela, *The isomorphism problem for hyperbolic groups, I*, Ann. of Math. (2) **141** (1995), n°2, 217–283.

[Sel97] Z. Sela, *Structure and rigidity in (Gromov) hyperbolic groups and discrete groups in rank* 1 *Lie groups, II*, GAFA, Geom. Funct. Anal. **7** (1997), n°3, 561–593.

[Sel99] Z. Sela, *Endomorphisms of hyperbolic groups, I: The Hopf property*, Topology **38** (1999), n°2, 301–321.

[Sel] Z. Sela, *Diophantine geometry over groups, VI: The elementary theory of a free group*, to appear in GAFA, Geom. Funct. Anal.

[Ser77] J.-P. Serre, *Arbres, amalgames,* SL_2, Astérisque **46**, Soc. Math. Fr. (1997).

[Sha00] Y. Shalom, *Rigidity of commensurators and irreducible lattices*, Invent. Math. **141** (2000), n°1, 1–54.

[Sho91a] H. Short et al., *Notes on word hyperbolic groups*, in *Group theory from a geometrical viewpoint*, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991), 3–63.

[Sho91b] H. Short, *Quasiconvexity and a theorem of Howson's*, in *Group theory from a geometrical viewpoint*, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991), 168–176.

[Shu99] A. G. Shukhov, *On the dependence of the growth exponent on the length of the defining relation*, Math. Notes **65** (1999), n°3–4, 510–515.

[Sil03] L. Silberman, *Addendum to: "Random walk in random groups"*, GAFA, Geom. Funct. Anal. **13** (2003), n°1, 147–177.

[Ska99] G. Skandalis, *Progrès récents sur la conjecture de Baum-Connes. Contribution de Vincent Lafforgue*, séminaire Bourbaki **829** (1999). Published in Astérisque **276** (2002), 105–135.

[Sta83] J. Stallings, *Topology of finite graphs*, Invent. Math. **71** (1983), n°3, 551–565.

[Sti82] J. Stillwell, *The word problem and the isomorphism problem for groups*, Bull. Amer. Math. Soc. (N.S.) **6** (1982), n°1, 33–56.

[STY02] G. Skandalis, J.-L. Tu, G. Yu, *The coarse Baum-Connes conjecture and groupoids*, Topology **41** (2002), n°4, 807–834.

[Swa96] G. A. Swarup, *On the cut point conjecture*, Electron. Res. Announc. Amer. Math. Soc. **2** (1996), n°2, 98–100 (electronic).

[TV99] S. Thomas, B. Velickovic, *On the complexity of the isomorphism relation for finitely generated groups*, J. Algebra **217** (1999), n°1, 352–373.

[Val02a] A. Valette, *Nouvelles approches de la propriété* (T) *de Kazhdan*, séminaire Bourbaki **913** (2002). Published in Astérisque **294** (2004), 97–124.

[Val02b] A. Valette, *Introduction to the Baum-Connes conjecture*, from notes taken by Indira Chatterji, with an appendix by Guido Mislin, Lectures in Mathematics ETH Zürich, Birkhäuser (2002).

[Ver00] A. M. Vershik, *Dynamic theory of growth in groups: entropy, boundaries, examples*, Russian Math. Surveys **55** (2000), n°4, 667–733.

[Wan98] M.-T. Wang, *A fixed point theorem of discrete group actions on Riemannian manifolds*, J. Differential Geom. **50** (1998), n°2, 249–267.

[Wis04] D. T. Wise, *Cubulating small cancellation groups*, GAFA, Geom. Funct. Anal. **14** (2004), n°1, 150–214.

[Wis] D. T. Wise, *Sixtolic complexes and their fundamental groups*, preprint.

[Woe00] W. Woess, *Random walks on infinite graphs and groups*, Cambridge Tracts in Mathematics **138**, Cambridge University Press (2000).

[Yu00] G. Yu, *The coarse Baum-Connes conjecture for spaces which admit a uniform embedding into Hilbert space*, Invent. Math. **139** (2000), n°1, 201–240.

[Żuk96] A. Żuk, *La propriété (T) de Kazhdan pour les groupes agissant sur les polyèdres*, C. R. Acad. Sci. Paris Sér. I Math. **323** (1996), n°5, 453–458.

[Żuk03] A. Żuk, *Property (T) and Kazhdan constants for discrete groups*, GAFA, Geom. Funct. Anal. **13** (2003), n°3, 643–670.

Sharp phase transition theorems for hyperbolicity of random groups

*Ce texte contient les principaux résultats de ma thèse. Il est reproduit ici car il sert de base à plusieurs travaux ultérieurs. Il a été publié dans le volume 14, n°3 (2004) de GAFA (*Geometry and Functional Analysis*), pp. 595–679. Les résultats avaient été préalablement annoncés par la note* Critical densities for random quotients of hyperbolic groups *publiée au volume 336, n°5 (2003) des* Comptes Rendus – Mathématique *de l'Académie des sciences de Paris, pp. 391–394.*

Sharp phase transition theorems for hyperbolicity of random groups

Yann Ollivier

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.

Contents

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'shanskiĭ in [Ols1], and independently by C. Champetier in [Ch1]: fix m and N and consider the group G presented by $\langle a_1, \dots, a_m; r_1, \dots r_N \rangle$ 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 larger.

This model goes as follows: Choose at random N cyclically reduced words of length ℓ in the letters $a_i^{\pm 1}$, uniformly among the set of all such cyclically reduced words (recall a word is called *reduced* if it does not contain a sequence of the form $a_i a_i^{-1}$ or $a_i^{-1} a_i$ and *cyclically reduced* if moreover the last letter is not the inverse of the first one). 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 cyclically 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 is 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 interpret $d\ell$ as 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 (with "harmless" torsion), 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.

We need a technical assumption of "harmless" torsion (see Definition 11). Hyperbolic groups with harmless torsion include torsion-free groups, free products of torsion-free groups and/or finite groups (such as $\mathrm{PSL}_2(\mathbb{Z})$), etc. This assumption is necessary: Indeed there exist some hyperbolic groups with non-harmless torsion for which Theorem 4 does not hold. [1]

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, or cyclically reduced words (the theorem is identical for these two variants).

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 with harmless torsion, 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 (cyclically) 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.

[1] These results were announced in [Oll1] without this assumption. I would like to thank Prof. A.Yu. Ol'shanskiĭ for having pointed an error in the first version of this manuscript regarding the treatment of torsion, which led to this assumption and to counterexamples to be presented in a forthcoming paper.

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 with harmless torsion, 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 with harmless torsion, 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 $(2m)^{d\ell}$ 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-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'shanskiĭ 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, easily satisfies 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.

In [Z], A. Żuk proves that a random quotient of the free group by reduced words at density greater than 1/3 has property T. As a random quotient of any group is the quotient of a random quotient of the free group by the relations defining the initial group, this means that the random quotients we consider possess property T as well for reduced words and densities above 1/3.

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, the fact that most subgroups are free, planarity of the Cayley graph, or the isomorphism problem; and more are to come. See for example [Ch1], [AO], [A], [Z], [AC], [KS].

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

The use of generic properties of groups also led to an announcement of an enumeration of one-relator groups up to isomorphism, see [KSS].

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.

Special thanks to Prof. A.Yu. Ol'shanksiĭ, who kindly pointed out an error in the treatment of torsion in a previous version of the manuscript (the assumption of harmless torsion did not appear), which led to counterexamples to be presented in a forthcoming paper, as well as for careful reading and suggestions for the text.

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 in receiving me.

1 Definitions and notation

1.1 Basics

Throughout all this text, G will be a discrete hyperbolic group given by a presentation $\langle S; R\rangle$ where $S = \{a_1, \ldots, a_m, a_1^{-1}, \ldots a_m^{-1}\}$ 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 $a \in S$ immediately 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_G^ℓ 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 g controls the asymptotics of the number of elements of S_G^ℓ: this number is roughly equal to $(2m)^{g\ell}$. The gross cogrowth θ 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 η 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 $a_1, \ldots, a_m$ is defined as*

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

The gross cogrowth *of the group G with respect to the generating set $a_1, \ldots, a_m$ 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 $a_1, \ldots, a_m$ 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. All of them are 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.

Alternative 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 $g/2 + \theta \geqslant 1$.

Gross cogrowth and cogrowth are linked by the following equation (see [Gri]):

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

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 (see [K1]). 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 formula 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.)

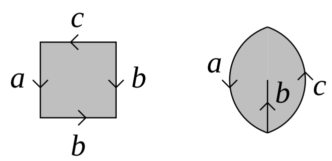
The cogrowth and gross cogrowth depend on the generating set. For example, adding trivial generators $a_i = e$ makes them arbitrarily close to 1.

1.3 Diagrams

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

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

A *decorated abstract van Kampen diagram* (davKd for short) is defined almost the same way as a van Kampen 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 and 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 davKd 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 van Kampen diagram with respect to presentation $\langle S; R\rangle$.

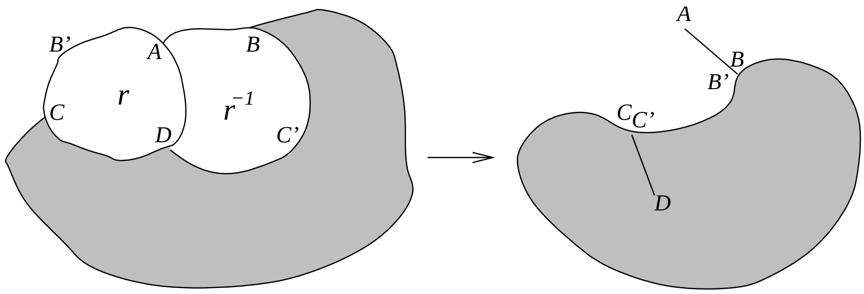
A *davKd with border* $w_1, \ldots, w_n$, where $w_1, \ldots, w_n$ are words, will be a $(n-1)$-hole davKd with each boundary edge decorated by a letter such that the words read on the n components of the boundary are $w_1, \ldots, w_n$. A davKd with border is said to be *fulfillable* if, as a davKd, 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, maybe filamenteous, davKd with border w is fulfillable (see [LS]).

A van Kampen 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 in the relator (w.r.t. the starting point). A davKd 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 in the relator.

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

Note that a minimal van Kampen 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 "van Kampen diagram or fulfillable decorated abstract van Kampen diagram". We will use the term *minimal diagram* as a short-hand for "minimal van Kampen diagram or minimal fulfillable decorated abstract van Kampen diagram with border".

1.4 Isoperimetry and narrowness

There is a canonical metric on the 1-skeleton of a van Kampen diagram (or a davKd), 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 $C > 0$ 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 proceed here 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 inverse of 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 greater than $1/2$, this will not occur only 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 large 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 $\langle (a_i)\ ;\ a_i = a_i^{\pm 1}, a_i = a_j\ \forall i, j \rangle$ 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 $d < 1/2$

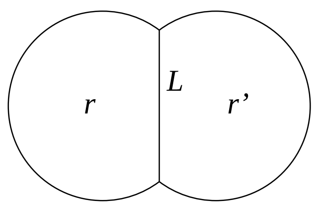
We proceed as follows: We will show that the only (reduced) davKd's which are fulfillable by a random presentation are those which satisfy some linear isoperimetric inequality. This is stronger than proving that only minimal diagrams satisfy an isoperimetric inequality: in fact, *all* reduced diagrams 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 on a chessboard.)

Thus we will evaluate the probability that a given decorated abstract van Kampen diagram can be fulfilled by a random presentation. We show that if the davKd 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 davKd's with at most K faces, violating the isoperimetric inequality. The probability that at least one of them is fulfillable is less that $D(K)$ times some quantity decreasing exponentially with ℓ, and taking ℓ large enough ensures that with probability arbitrarily close to one, none of these davKd's is fulfillable. The conclusion then follows by the Cartan-Hadamard-Gromov-Papasoglu theorem.

The basic picture is as follows: Consider a davKd 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 except 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 greater 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 davKd. 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 of each face $1, 2, \ldots, \ell$ starting at the marked point, following the given orientation.

If, in the davKd 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.

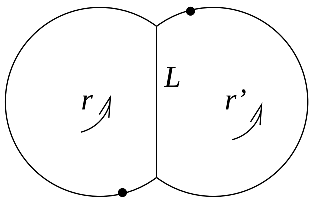
Thus, a -1 edge between the k-th vertex of the i-th part and the k'-th vertex of the j-th part means that the k-th letter of relator r_i has to be the inverse of the k'-th letter of relator r_j.

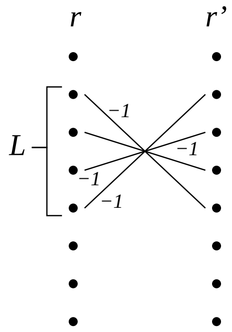
Successively add an edge to Γ in this way for each interior edge of the davKd D, so that the total number of edges of Γ is equal to the number of interior edges of D.

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

Note that this graph only depends on the davKd 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 largest 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's are 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 reduced davKd, 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 davKd violating the isoperimetric inequality tends to 0 when $\ell \to \infty$. But here we use the local-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 van Kampen diagram with at most K faces satifies

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

then any reduced van Kampen 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 davKd'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 davKd 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 davKd it represents, and with vertices of degree at least 3 (and, as in a davKd, 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. We can of course use the general argument of section 6.9.1 but there is a shorter proof in this case. First, as any reduced diagram satisfies $|\partial D| \geqslant \alpha'\ell\,|D| \geqslant \alpha'\ell$, the ball of radius $\alpha'\ell/2$ injects into the quotient, hence the quotient contains at least one non-trivial element and cannot be $\{e\}$.

Second, we prove that the presentation is aspherical. With our conventions on van Kampen diagrams, our asphericity implies asphericity of the Cayley complex and thus cohomological dimension at most 2 (indeed, thanks to the marking of each face by a starting point and a relator number, two faces are reducible in a diagram only if they really are the same face in the Cayley complex, so that diagram reduction is a homotopy in the Cayley complex). This will end the proof: indeed, cohomological dimension at most 2 implies torsion-freeness (see [B], p. 187), hence the quotient cannot be a non-trivial finite group.

Indeed, the isoperimetric inequality above is not only valid for minimal diagrams, but for *any* reduced diagram. Now suppose that there is some reduced spherical diagram. It will have zero boundary length and thus will violate any isoperimetric inequality, hence a contradiction. Thus the presentation is aspherical.

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.

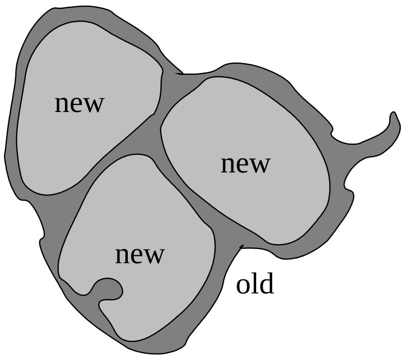
We will give a general theorem for hyperbolicity of random quotients by words taken from some probability measures on the set of all words. We will need somewhat technical axioms on the measures (for example, that they weight only long words). Here we give a heuristic justification of why these axioms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

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

If D is a van Kampen 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").

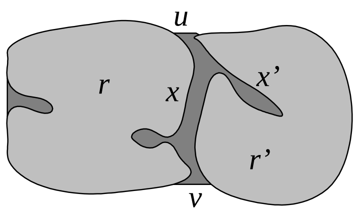
Say the new relators have length of order ℓ where ℓ is much larger than the hyperbolicity constant of G. (This will be Axiom 1.)

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 $(2m-1)^{-L}$. But we now have to consider the case when 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 $xx'^{-1} = e$; but xx'^{-1} is a uniformly chosen random word of length $L+L'$, 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 alternative 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 parts 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 size of centralizers of torsion elements in 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 greater than 2ℓ. This is exactly what we want to prove: either the boundary is long, 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. Similarly, 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 the measure satisfies some simple axioms, then the random quotient by elements picked under the measure is hyperbolic.

For each of the elements of G weighted by 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 notation

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.

Accordingly, we will say that $f(\ell, L) \approx g(\ell, L)$ uniformly for all $L \leqslant \ell$ if whatever the sequence $L(\ell) \leqslant \ell$ is, we have

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

and if this limit is uniform in the sequence $L(\ell)$.

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 $a, b \in [0; 1]$, $b > 0$, whenever we pick a word x according to μ_ℓ we have

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

or

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

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).

We also have to condition w.r.t. the length of the word since in the definition of a random set of density d under μ_ℓ above, we made a sampling for each length separately.

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 $b > 0, b' > 0, a+b \leqslant 1, a'+b' \leqslant 1$, whenever we pick two independent words x, x' according to μ_ℓ we have

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

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

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

We give similar definitions when $a+b > 1$ or $a'+b' > 1$, conditioning by every subword not in $x_{a;b}$ or $x'_{a';b'}$.

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-global principle (Appendix A).

Axiom 1.
There is a constant $\kappa_1 \geqslant 1$ 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 not 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 probably not 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 $t \leqslant 1$, we have

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

uniformly in t.

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_3 such that for any two disjoint subwords x, y under μ_ℓ, for any $n = n(\ell)$, the probability that there exist words u and v of length at most n, such that $xuyv = e$ in G, is roughly less than $(2m)^{\gamma_3 n}(2m)^{-\beta_3(|x|+|y|)}$.

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

Axiom 4.
There exist constants β_4 and γ_4 such that, for any subword x under μ_ℓ, for any $n = n(\ell)$, the probability that there exist words u and v of length at most n, such that $ux = xv$ and $u \neq e$, $v \neq e$, is roughly less than $(2m)^{\gamma_4 n}(2m)^{-\beta_4 |x|}$

If G has large centralizers, this axiom will probably fail to be true. We will see below (section 4.5) that, in a hyperbolic group with "strongly harmless" torsion, 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 $\beta_{4'}$ and $\gamma_{4'}$ such that, for any $C > 0$, for any subword x under μ_ℓ, for any $n = n(\ell)$, the probability that there exists a word u of length at most n such that some cyclic permutation x' of xu satisfies $\|x'\| \leqslant C \log \ell$, is roughly less than $(2m)^{\gamma_{4'} n}(2m)^{-\beta_{4'}|x|}$.

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. 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 some family of measures μ_ℓ, it applies as well to any family of measures $\mu'_\ell \lesssim \mu_\ell$.

4.5 On torsion and Axiom 4

We show here that in a hyperbolic group with "harmless" torsion, 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 elements commuting with torsion elements, nor close to powers of short elements.

Recall that the virtual centre of a hyperbolic group is the set of elements whose action on the boundary at infinity is trivial. For basic properties see [Ols2].

Definition 11 (Harmless torsion).

A torsion element in a hyperbolic group is said to be strongly harmless *if its centralizer is either finite or virtually $\mathbb{Z}$.*

A torsion element is said to be harmless *if it is either strongly harmless or lying in the virtual centre.*

A hyperbolic group is said to be with (strongly) harmless torsion *if each non-trivial torsion element is (strongly) harmless.*

Harmfulness is defined as the opposite of harmlessness.

For example, torsion-free groups are with harmless torsion, as well as free products of free groups and finite groups. Strongly harmless torsion is stable by free product, but harmless torsion is not.

Let μ_ℓ be a measure satisfying Axioms 1, 3 and 4'.

Proposition 12.
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 $(2m)^{\gamma_{4'} n}(2m)^{-\beta_{4'}|x|}$.

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 50 (Appendix B) 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 $(2m)^{\gamma_{4'} n}(2m)^{-\beta_{4'}|x|}$. □

Proposition 13.
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 $p \lesssim (2m)^{-\beta_3|x|}$. □

Proposition 14.
Suppose G has strongly harmless torsion, and that Axioms 1, 3 and 4' are satisfied. Set $\beta = \min(\beta_3, \beta_{4'})$.

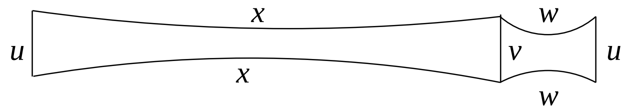
There is a constant γ such that for any subword x under μ_ℓ, the probability that there exist words u, v of length at most $n = n(\ell)$, such that $ux = xv$ in G, with u, v not equal to e, is roughly less than $(2m)^{\gamma n - \beta|x|}$.

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 50 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 $(2m)^{\gamma_{4'}Cn-\beta|x|}$ by Proposition 12. In this case we conclude.

Now suppose that xw is not a torsion element. Then we can glue the above diagram to copies of itself along their u-sides. This way we get two quasi-geodesics labelled by $((xw)^n)_{n\in\mathbb{Z}}$ that stay at a 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 a hyperbolic element, or (by strong harmlessness) that u is a torsion element with virtually cyclic centralizer.

The idea is that in this situation, xw will lie close to some geodesic Δ depending only on the short element u. As there are not many such Δ's (and as the probability for a random word to be close to a given geodesic behaves roughly like the probability to be close to the origin), this will be unlikely.

First, suppose that u is hyperbolic. 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), and set $\Delta' = xw\Delta$. As xw stabilizes the limit of Δ, Δ' is equivalent to Δ. But two equivalent geodesics in a hyperbolic group stay at Hausdorff distance at most R_1 where R_1 is a constant depending only on the group (see [GH], p. 119).

The distance from xw to Δ' is equal to the distance from e to Δ. By Proposition 51 applied to $u^0 = e$, this distance is at most $|u| + R_2$ where R_2 is a constant depending only on G. Hence the distance from xw to Δ is at most $|u| + R$ with $R = R_1 + R_2$. Let y be a point on Δ realizing this distance. As $|xw| \leqslant |x| + |w|$, we have $|y| \leqslant |x| + |w| + |u| + R$. There are at most $2|x| + 2|w| + 2|u| + 2R + 1$ such possible points on Δ (since Δ is a geodesic). For each of these points, the probability that x falls within distance $|u| + R + |w|$ of it is roughly less than $(2m)^{|u|+R+|w|}(2m)^{-\beta|x|}$ by Proposition 13 applied to all of these points. So the probability that x falls within distance less than $|u|+R+|w|$ of any one of the possible y's on a given geodesic Δ is roughly less than $(2|x|+2|w|+2|u|+2R+1)(2m)^{|u|+R+|w|}(2m)^{-\beta|x|}$ which in turn is roughly less than $(2m)^{Cn-\beta|x|}$ as $|w| \leqslant Cn$ and R is a constant.

This was for one fixed u. 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. Thus we can conclude when u is hyperbolic.

Second, if u is a torsion element with virtually cyclic centralizer Z, we use a similar argument. Let L as above be the maximal length of a shortest element of a conjucacy class of a torsion element. By Proposition 49, u is conjugate to some torsion element u' of length at most L by a conjugating word v with $|v| \leqslant |u|/2 + R_1$ where R_1 is a constant. The centralizer of u' is $Z' = vZv^{-1}$. We know that $xw \in Z$.

There are two subcases: either Z is finite or Z is virtually $\mathbb{Z}$.

Let us begin with the former. If Z is finite, let $\|Z\|$ be the maximal norm of an element in Z. We have $\|Z\| \leqslant 2|v| + \|Z'\|$. Let $R_2 = \max \|Z'\|$ when u' runs through all torsion elements of norm at most L. As xw lies in Z we have $\|x\| \leqslant |w| + \|Z\| \leqslant |w| + 2|v| + R_2 \leqslant |w| + |u| + 2R_1 + R_2$. So by Proposition 13 the probability of this event is roughly less than $(2m)^{|w|+|u|+2R_1+R_2} \lesssim (2m)^{Cn+n}$ as $|w| \leqslant Cn$ and as R_1, R_2 are mere constants.

Now if Z is virtually $\mathbb{Z}$, let Δ be a geodesic joining the two limit points of Z. The element u' defined above stabilizes the endpoints of the geodesic $v\Delta$, and so does $vxwv^{-1}$.

By Corollary 53, $vxwv^{-1}$ lies at distance at most $R(v\Delta)$ from $v\Delta$. As there are only a finite number of torsion elements u' with $\|u'\| \leqslant L$, the supremum R of the associated $R(v\Delta)$ is finite, and so, independently of u, the distance between $vxwv^{-1}$ and $v\Delta$ is at most R.

Now $\operatorname{dist}(xw, \Delta) \leqslant |v| + \operatorname{dist}(xwv^{-1}, \Delta) = |v| + \operatorname{dist}(vxwv^{-1}, v\Delta) \leqslant |v| + R$ and we conclude exactly as in the case when u was hyperbolic, using that $|v| \leqslant |u|/2 + R_1$. This ends the proof in case u is a torsion element with virtually cyclic centralizer. □

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 strongly harmless torsion (instead of harmless torsion), 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 into 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 from 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 alternative 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 15.
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 about 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 16.
The probability that B_ℓ is equal to e is roughly less than $(2m)^{-(1-\theta)\ell}$.

Proof.
Alternate definition. □

Proposition 17.

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

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 18.
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 19.
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 17 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 20.
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 words. 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 alternative 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 $x = x_1 x_2 \ldots x_N$ 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 notation 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 two elements $x, y \in G$. That is, $(x|y) = \frac{1}{2}\left(\|x\| + \|y\| - \|x^{-1}y\|\right)$.

Proposition 21.
Let $x \in G$ and $L \leqslant \ell$. We have

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

uniformly in x and $L \leqslant \ell$.

Proof.
Let y be the point at distance L on a geodesic joining e to x. By the triangle-tripod transformation in exX_ℓ, the inequality $(x|X_\ell) \geqslant L$ 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. □

Corollary 22.
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 x and 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 21 to X_ℓ and x^{-1} and write out the definition of the Gromov product. □

Proposition 23.
For any fixed N, uniformly for any $x\in G$ and 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 22, 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. □

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 to prove 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 $|x|/\ell$ of the subword is correct since we did not ask the estimates to be uniform in this ratio.

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 24.
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 23 starting at e, this occurs with probability $(2m)^{-g(|x|-L)/2}$. Now we have $\|xu\|\geqslant L-\|u\|$, but $\|xuy\|=\left\|v^{-1}\right\|$. By Proposition 23 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 25.
Axiom 4' is satisfied for random geodesic words, with exponent $g/2$.

Proof.
Taking notation 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 23 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 23 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 $(2m-1)^{\eta\ell}$.

Here we have to suppose $m > 1$. (A random quotient of $\mathbb{Z}$ by reduced words of length ℓ is $\mathbb{Z}/\ell\mathbb{Z}$.)

Proposition 26.
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 (if $m > 1$).

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 greater than $1/2$ anyway.

In particular, 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 $S = a_1^{\pm 1}, \dots, a_m^{\pm 1}$. Let R be a set of $(2m)^{d\ell}$ randomly chosen words of length ℓ. We 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 greater than the ambient dimension do intersect (cf. our discussion of the density 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 (with $m > 1$ in the reduced word model and $g > 0$ in the geodesic word model) 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 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 $(2m)^{g\ell}$, with $g > 0$ as G is non-elementary.

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 in order to have the number of words taken go to infinity, we have to suppose that $m \geqslant 2$.

5.3 Elimination of the virtual centre

Theorem 9 only applies to random quotients of hyperbolic groups with strongly harmless torsion. We have to show that the presence of a virtual centre does not change random quotients. 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 [Ols2] or [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 27.
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 28.
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 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 van Kampen 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 van Kampen 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.

6.1 On the lengths of the relators

In order not to make the already complex notation 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.

For the general case, there are only three ways in which the length of the elements matters for the proof:

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

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

3. In order not to perturb our probability estimates, the number of distinct lengths of the relators in R must be subexponential in ℓ.

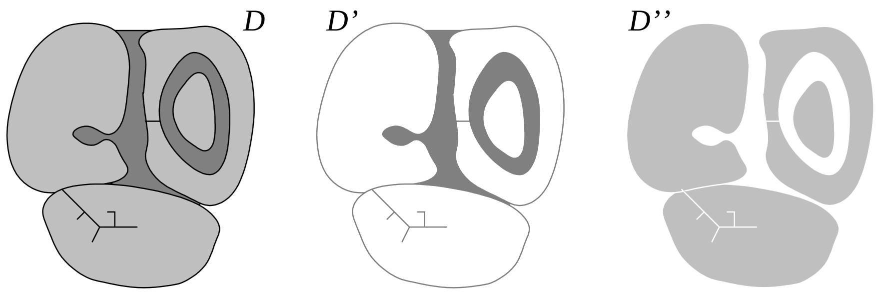
All these properties are guaranteed by Axiom 1.

6.2 Combinatorics of van Kampen 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 van Kampen 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'.



After this manipulation, we consider that each edge of D'' is in contact only with an edge of D', so that we never have to deal with equalities between subwords of two new relators (we will treat them as two equalities to the same word — cf. the definition of coarsening below).

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 slightly stronger meaning than previously, which we define now.

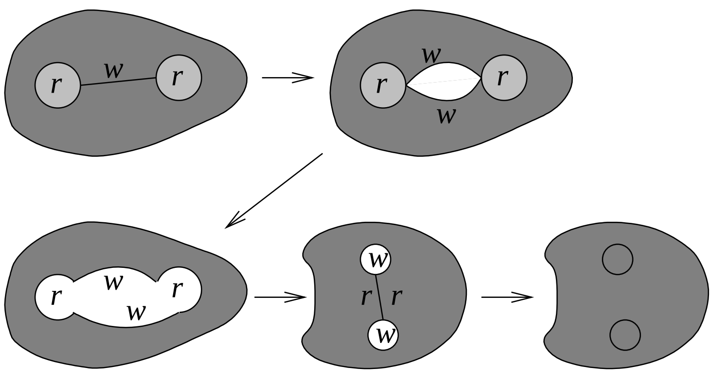
Definition 29.
A van Kampen diagram $D = D' \cup D''$ on $G/\langle R \rangle$ (with D' and D'' as above) is said to be strongly reduced *with respect to G if there is no pair of faces of D'' bearing the same relator with opposite orientations, such that their marked starting points are joined in D' by a simple path representing the trivial element in G.*

In particular, a strongly reduced diagram is reduced.

Proposition 30.
Every van Kampen 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 van Kampen diagram D such that D' is minimal very probably satisfies some linear isoperimetric inequality. By the local-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 31.
There exist constants $\alpha, \alpha' > 0$ (depending on G and d but not on ℓ) such that, for any integer K, with probability exponentially close to 1 as $\ell \to \infty$ the set of relators R satisfies the following:

For any van Kampen diagram $D = D' \cup D''$ satisfying the three conditions:

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

then D satisfies the isoperimetric inequality

$$|\partial D| \geqslant \alpha\ell \left|D''\right| + \alpha' \left|D'\right|$$

(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. This step uses the local-global hyperbolic principle (Appendix A), which essentially states that it is enough to check the isoperimetric inequality for a finite number of diagrams.

Proposition 32.
There exists an integer K (depending on G and d but not on ℓ) such that if the set of relators R happens to satisfy the conclusions of Proposition 31, with ℓ large enough, then $G/\langle R\rangle$ is hyperbolic. Better, then there exist constants $\alpha_1, \alpha_2 > 0$ such that for any strongly reduced diagram D such that D' is minimal, we have

$$|\partial D| \geqslant \alpha_1\ell \left|D''\right| + \alpha_2 \left|D'\right|$$

Remark 33.
Proposition 32 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 van Kampen diagram there exists another van Kampen 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 42. Take for property P in this proposition "to be strongly reduced". Recall the notation of Appendix A: $L_c(D) = |\partial D|$ 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 van Kampen 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 independent of ℓ to be chosen later. As $A_d(D) \leqslant K\ell^2$, we have $|D''| \leqslant K$. Proposition 31 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 42 are fulfilled by $K\ell^2$. (The important point is that this K is independent of ℓ.)

The conclusion is that any strongly reduced van Kampen 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

$$|\partial D| \geqslant \alpha_1\ell\left|D''\right| + \alpha_2\left|D'\right|$$

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

So the proposition above, combined with the local-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

$$|\partial D| \geqslant \alpha_1\ell\left|D''\right| + \alpha_2\left|D'\right|$$

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 31.

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 van Kampen 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 56 for diagrams with holes in G, we have that $|\partial D'| \geqslant C\,|D'| - |D''|\,\lambda(2+4\alpha\log|D'|)$. So, for ℓ large enough,

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

as for ℓ large 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* $|D'| \leqslant 4\,|D''|\,\ell/C$.

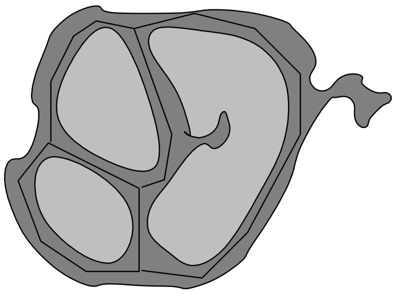
6.3 Coarsening of a van Kampen diagram

We now define the coarsening of a van Kampen diagram: this will be the van Kampen 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 van Kampen diagram with respect to the new relators, with some narrow "glue" (that is, old relators) between faces. (This "glue" has some similarity to "contiguity subdiagrams" in [Ols1].)

The diagram D' has at most K holes. After Corollary 57, 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.

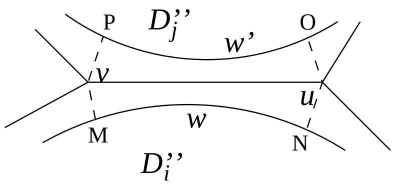
So D' is $E\log\ell$-narrow. This means that a point of D' is either $E\log\ell$-close to some point of D'' or to some point of the boundary of D.

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).

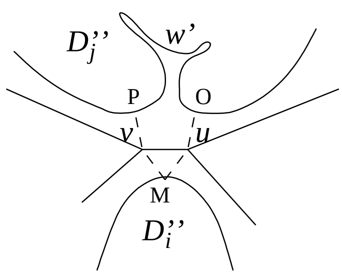


The D_i's for $1 \leqslant i \leqslant K$ form a planar graph X, which is a kind of van Kampen diagram at the scale of the new relators. Denote by D''_i the i-th face of D'', so that $D''_i \subset D_i$.

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 boundary of D''_i, w' lies on the boundary 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 davKd, except that each edge bears two lengths instead of one. *From now on, we redefine a davKd to be such a decorated graph.*

A davKd is said to be *fulfillable* if it is the coarsening of some *strongly reduced* van Kampen diagram D of $G/\langle R\rangle$. We have to show that any fulfillable davKd 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 van Kampen 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 davKd 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 davKd. 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) van Kampen 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.

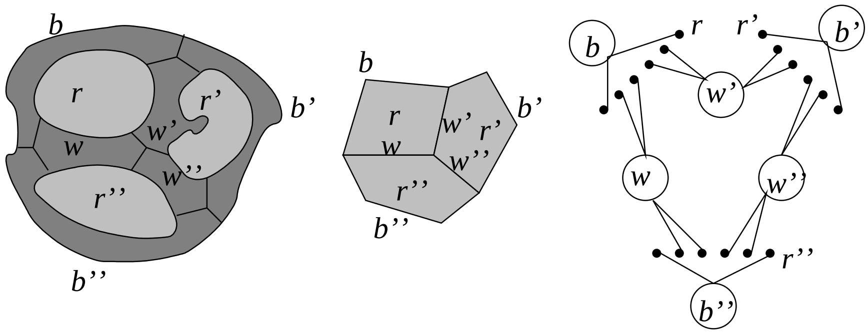
(This may result in double edges if $i = i'$; we will deal with this problem later.)

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 van Kampen 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 van Kampen diagram which is equal to e in G.

Indeed, fulfillability of the davKd 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 van Kampen diagram, two adjacent faces bear the same relator.

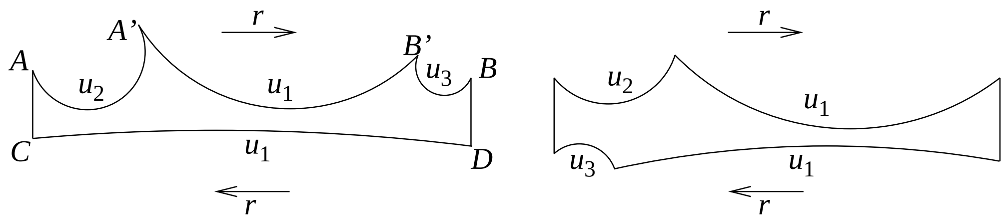
Doublets are annoying since the two sides of the translator are not chosen independently, whereas our argument requires some degree of independence. We will split the corresponding translators to control the occurrences of such a situation.

This section is only technical.

Consider a translator in the van Kampen 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.

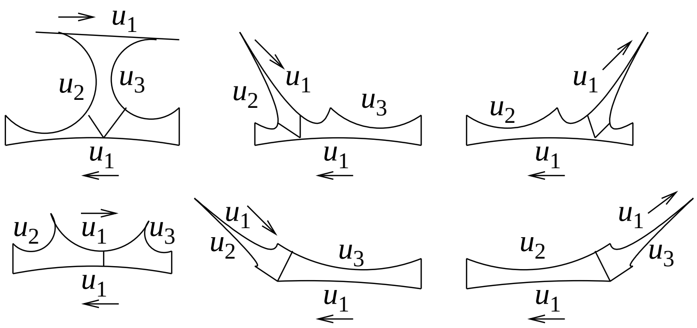
By elementary hyperbolic geometry, and given that the two lateral sides are of length at most $2E\log\ell$, any point on CD is $(2\delta+2E\log\ell)$-close to some point on AA' or $B'B$, or 2δ-close to some point on $A'B'$.

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 AB.

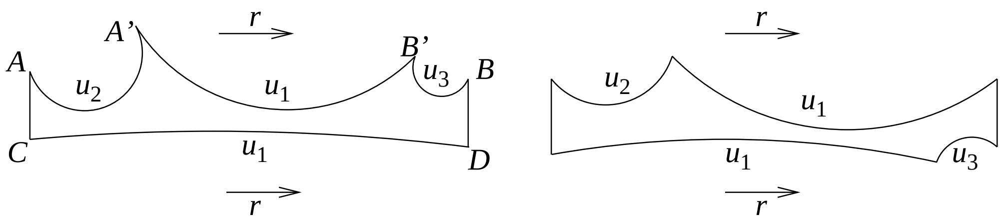
There are six cases depending on whether C'_{i_0} and C'_{i_0-1} belong to AA', $A'B'$ or $B'B$. In each of these cases we can cut the diagram in at most three parts, in such a way that no part contains two copies of some subword of u_1 (except perhaps up to small words of length at most 8δ at the extremities). 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 $u' = u_1u_3$.



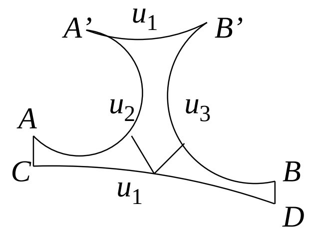
We will only treat the first case, as the second is similar.

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 $A'B'$.

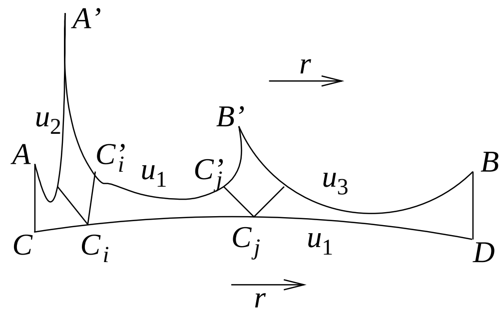
If any point on CD is close to a point on either AA' or BB', we can simply eliminate the doublets by cutting the figure at the last point of CD which is close to AA'. (As above, by cutting the figure we mean that we split the vertex of Γ representing the translator into three new vertices and we share 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 $\varphi(C_i) = \operatorname{dist}(C'_i, A')$ 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 into 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).



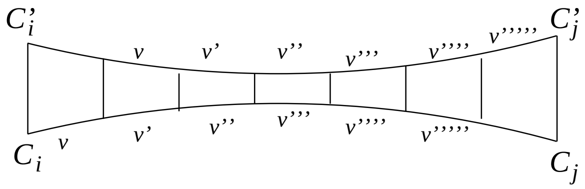
This way, we get a figure in which only the middle part $C_iC_jC'_jC'_i$ of the figure contains two copies of a given piece of u_1. Indeed (from left to right in the figure) the first part contains letters 0 to i of the lower copy of u_1 and no letter of the upper u_1; the second part contains letters 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 that 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 the graph 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 $2\varepsilon_2 |u_1|$ 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 $\varepsilon_1 + 2\varepsilon_2$ in each translator.

6.6 Pause

Let us sum up the work done so far. Remember the example on page 177.

Proposition 34.
For each strongly reduced van Kampen 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.
- The number of edges at any ordinary vertex is at most equal to the number of times the corresponding relator of R appears in D.
- 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 $\varepsilon_2 \left|u\right|$.
- 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 $\left|D''\right| \ell(1 - \varepsilon_1 - 2\varepsilon_2)$.

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 van Kampen 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 notation) in mind for the sequel.

6.7 Apparent length

The line of the main argument below is to fulfill the davKd 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 35 (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 $\varepsilon_3\ell/\kappa_1 \leqslant |y| \leqslant \kappa_1\ell$ with κ_1 as in Axiom 1, for some ε_3 to be fixed soon. We will also use the evaluation from Axiom 2.

Definition 36 (Apparent length).
The apparent length *of x is*

$$\mathbb{L}(x) = \min\left(\|x\|/\kappa_2, \min_{\varepsilon_3\ell/\kappa_1 \leqslant L \leqslant \kappa_1\ell} \mathbb{L}_L(x)\right)$$

Our main tool will now be the following

Proposition 37.
For any subword x under μ_ℓ, we have

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

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 thus 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 with $d < \beta$, subwords of apparent length greater than L only occur with probability exponentially small as $\ell \to \infty$. Thus, we can safely assume 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. In the geodesic word model, apparent length is simply the usual length. 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 enter the main step of the proof. We will consider a davKd and evaluate the probability that it is fulfillable. We will see that either the davKd satisfies some isoperimetric inequality, or this probability is very small (exponential in ℓ). It will then be enough to sum on all davKd's with at most K faces to prove Proposition 31.

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 obvious 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 35) 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 an internal translator both sides of which have length less than $\varepsilon_3\ell$ a *zero-sided translator*. 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 $q(t) \leqslant i < r(t)$. Say that translator t is not begun at time i if $i < p(t)$.

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 $L(b)$'s for all boundary translators b 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 b, the apparent length of $u(b)$ is $L(b)$.

(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 this 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 notation, we will neglect this problem.

Of course, fulfillability of the davKd 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 notation 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 37 (that is, Axioms 2 and 3) tells us that, conditionally 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 notation 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 notation 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 35) 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 van Kampen 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 $\varepsilon_1 + 3\varepsilon_2$ that was removed), counting commutation translators twice. Thus, in the sum above, each of the ℓ letters of a appears exactly m_a times, and so

$$A_a - A_{a-1} \geqslant m_a \left(\ell(1-\varepsilon_4-\varepsilon_5-\varepsilon_1-3\varepsilon_2) + \frac{\log_{2m} \mathrm{P}_{i_f} - \log_{2m} \mathrm{P}_{i_0-1}}{\beta} \right) + o(\ell) \quad (\star)$$

(Because of the removal of a proportion at most $\varepsilon_1 + 3\varepsilon_2$ 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 non-positive.

Setting $\varepsilon_6 = \varepsilon_4 + \varepsilon_5 + \varepsilon_1 + 3\varepsilon_2 + o(\ell)/\ell$ in $(\star)$ 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 146 (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 large).

Summing the inequalities above for a from 1 to n gives

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

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 36 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

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

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

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

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

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

By definition, the probability that the davKd 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

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

To put it in the exact form of Proposition 31, 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 davKd is fulfillable is less than $(2m)^{-\ell(\beta-d)/4}$, which decreases exponentially in ℓ.

In order to prove Proposition 31 we have not only to evaluate this probability for one davKd but for all davKd's having at most K faces. Note that a davKd 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 davKd violating the isoperimetric inequality decreases exponentially with ℓ.

This proves Proposition 31, 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

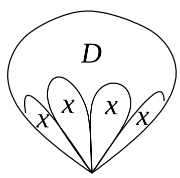
Let $d < \beta$. We will show that the probability that the random quotient is finite decreases exponentially as $\ell \to \infty$.

We know from hyperbolicity of the quotient (Proposition 32) that the probability that there exists a van Kampen diagram of the quotient whose part made of old relators is minimal 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 van Kampen diagram D bordered by x^n for some n.

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 above 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 van Kampen 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 van Kampen 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 in this family is fulfillable with respect to $R \cup \{x\}$.



By Proposition 30 we can take the strong reduction of this diagram. It is non-empty as the faces bearing x cannot be cancelled (they all have the same orientation).

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

But we know by what we already proved (Propositions 31 and 32) 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 large, d' is only slightly above d, and if ℓ is big enough we still have $d' < \beta$.

Say (from Proposition 32) 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, notation 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 van Kampen diagram of $G/\langle R\rangle$.

Now glue n faces containing y and p faces containing x to the boundary of this diagram. This defines a van Kampen diagram of $G/\langle R'\rangle$, which we can take the strong reduction of. This reduction is non-empty since faces bearing x and y cannot be cancelled (so in particular $|D''| \geqslant n+p$). 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 van Kampen 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 van Kampen diagrams is fulfillable by a strongly reduced van Kampen diagram on the relators of R'. But all these diagrams violate the isoperimetric inequality, hence the conclusion.

Alternative proof. We give an alternative 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 38.
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 $a_1^{\pm 1}, \dots a_m^{\pm 1}$. 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

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

by the central limit theorem with $2m-1$ degrees of freedom or by a direct computation using Stirling's formula.

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 $a_i = a_j^{\pm 1}$ 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 $\mathbb{Z}/2\mathbb{Z}$. □

Corollary 39.
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 dihedral 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 local-global principle, or Cartan-Hadamard-Gromov-Papasoglu theorem

The Cartan-Hadamard-Gromov-Papasoglu theorem allows us 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 drawn 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 40 (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 us to deduce global hyperbolicity properties. This was termed by Gromov *hyperbolic Cartan-Hadamard theorem* or *local-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 proves a somewhat stronger theorem.

Theorem 41 (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 42.
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

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

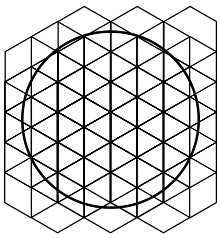
Then any disk D drawn in X, having P, with $A(D) \geqslant K^2$, satisfies

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

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 $L_c(B) \geqslant A_c(B)/10^4 K$ 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 which come from a disk drawn in X, and those where 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 43.
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 44.
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 45.
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 44, 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 amputating 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 43 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 45, 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 positive number such that any simply connected minimal van Kampen 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 cyclically geodesic *if it and all of its cyclic permutations label geodesic words in G.*

The following is well-known (cf. [BH], p. 452, where the authors use "fully reduced" for "cyclically geodesic").

Proposition 46.
Let u, v be cyclically geodesic 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 $2\delta + 1$.*

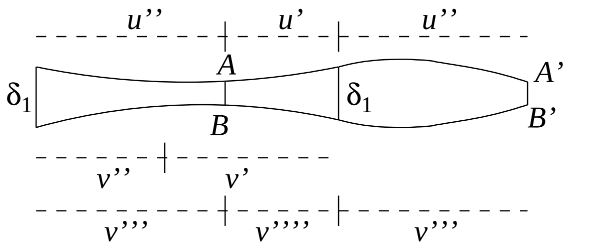
This immediately extends to:

Proposition 47.
Let u, v be cyclically geodesic 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 46. Construct the quadrilateral $u''u'\delta_1 v'^{-1} v''^{-1} \delta_1^{-1}$. As u and v are cyclically geodesic, 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 $v''v'$.



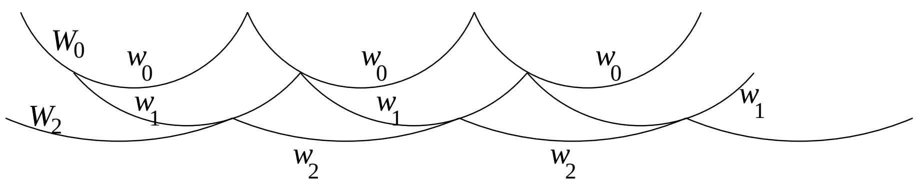
Let A be the endpoint of u''. The point A is $(2\delta + |\delta_1|)$-close to some point B on $v''v'$. Let δ_2 be a path connecting A to B. The point B divides $v''v'$ into two words v''' and v'''', and we have $u = u'u'' = \delta_2 v'''' v''' \delta_2^{-1}$ which ends the proof of the proposition.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

We will need the following

Proposition 48.
Let w be a geodesic word. There exists a cyclically geodesic word z which is conjugate to w by a word of length at most $(|w| - |z|)(\delta + 1/2) + 4\delta$.

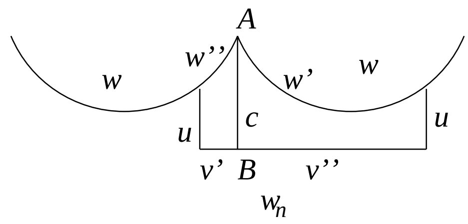
Proof.
Set $w_0 = w$ and construct a sequence w_n of geodesic words by induction. If w_n is cyclically geodesic, stop. If not, then write $w_n = w'_n w''_n$ such that $w''_n w'_n$ is not geodesic. Then set w_{n+1} to a geodesic word equal to $w''_n w'_n$. As length decreases at least by 1 at each step, the process stops after a finite number n of steps and w_n is cyclically geodesic. 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 at most $(|w''|+|w'|+2|u|-|w_n|)/2+4\delta$. 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. □

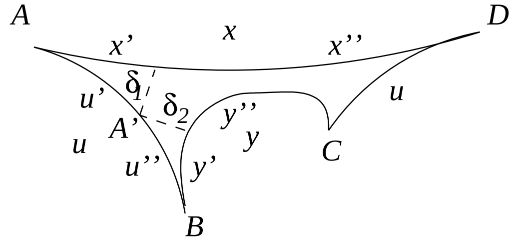
Now, in the spirit of Proposition 46, let $C_c = \max_{x,y} \min\{|u|, x = uyu^{-1}\}$ where the range of the maximum is the set of all couples of conjugate words of length at most $8\delta + 1$. As this set is finite we have $C_c < \infty$. Let $C'_c = C_c + 4\delta^2 + 12\delta + 2$.

Proposition 49.
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 van Kampen 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.

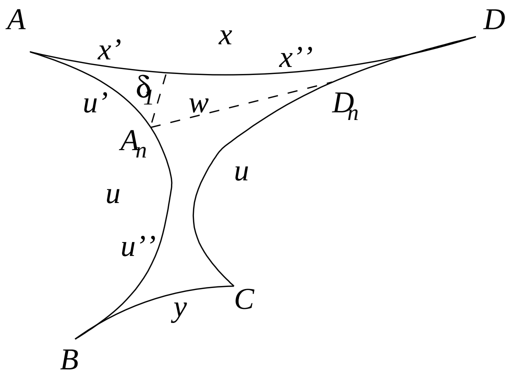


Let A' be the first point on AB which is 2δ-close to BC. Considering the point just before A', we know that A' is $(2\delta+1)$-close to AD.

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 at most $2\delta+1$.

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 cyclically geodesic. If w is cyclically geodesic as well, then we conclude by Proposition 47. If not, use Proposition 48 to find a cyclically geodesic 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 cyclically geodesic and we conclude by Proposition 47. □

Corollary 50.
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 a 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 Cyclic subgroups

We will also need the following.

Proposition 51.
There exists a constant R such that, for all hyperbolic $u\in G$, the Hausdorff distance between the set $(u^n)_{n\in\mathbb{Z}}$ and any geodesic with the same limit points is at most $\|u\|+R$.

Proof.

Lemma 52.
The Hausdorff distance between $(u^n)_{n\in\mathbb{Z}}$ and any geodesic with the same limit points is finite.

Proof of the lemma.
From [GH] (p. 150) we know that $k\mapsto(u^k)_{k\in\mathbb{Z}}$ is a quasi-geodesic. From [GH] (p. 101) we thus know that this quasi-geodesic lies at finite Hausdorff distance from some geodesic. From [GH] (p. 119) we know that any two geodesics with the same limit points lie at finite Hausdorff distance. □

Now for the proposition. First, suppose that u is cyclically geodesic. Let p be a geodesic path joining e to u. Let Δ be the union of the paths $u^n p$, $n\in\mathbb{Z}$. Since u is cyclically geodesic, Δ is a $(1,0,\|u\|)$-local quasi-geodesic (notation as in [GH]). Thus, there exist constants R and L depending only on G such that, if $\|u\|\geqslant L$, then Δ lies at Hausdorff distance at most R of some geodesic Δ' equivalent to it (see [GH], p. 101), hence at Hausdorff distance $16\delta+R$ of any other equivalent geodesic ([GH], p. 119). As there are only a finite number of u's such that $\|u\|<L$, and as for each of them the lemma states that Δ lies at finite Hausdorff distance from any equivalent geodesic, we are done when u is cyclically geodesic.

If u is not cyclically geodesic, apply Proposition 49 to get a cyclically geodesic word v such that $v=xu''u'x^{-1}$ with $u=u'u''$ and $|x|\leqslant C'_c$. Apply the above to $(v^k)_{k\in\mathbb{Z}}$: this set stays at distance at most R of some geodesic Δ. Translate by $u'x^{-1}$. The set $(u'x^{-1}v^k)_{k\in\mathbb{Z}}$ stays at distance at most R of the geodesic $u'x^{-1}\Delta$. But since $u^k=u'x^{-1}v^kxu'^{-1}$, the Hausdorff distance between the sets $(u^k)_{k\in\mathbb{Z}}$ and $(u'x^{-1}v^k)_{k\in\mathbb{Z}}$ is at most $\left\|xu'^{-1}\right\|\leqslant C'_c+\|u\|$ and we are done. □

Since the stabilizer of any point of the boundary is either finite or has $\mathbb{Z}$ as a finite index subgroup (cf [GH], p. 154), we get as an immediate by-product of the lemma

Corollary 53.
Let Δ be a geodesic in G, with limit points a and b. There exists a constant $R(\Delta)$ such that for any x in the stabilizer of a and b, the distance from x to Δ is at most $R(\Delta)$.

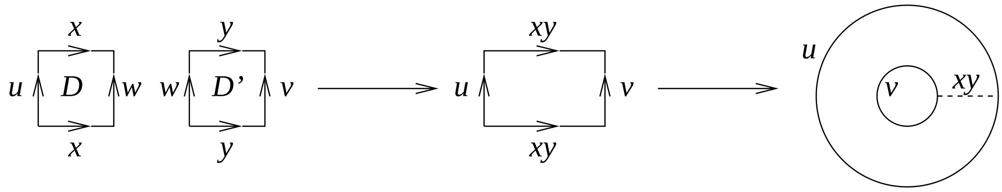
B.3 One-hole diagrams

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

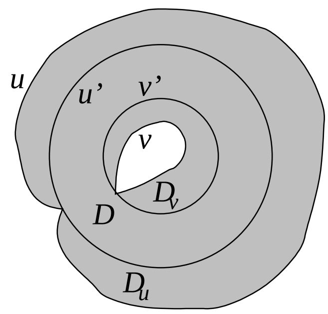
Proposition 54.
There exists a constant $C' > 0$ such that for any two conjugate words u and v, there exists a one-hole diagram D bordered by u and v, such that $C' |D| \leqslant |u| + |v|$.

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 49, u and w are conjugate by a word x of length at most $|u|/2 + |w|/2 + C'_c$. Thus, there exists a minimal van Kampen 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 van Kampen 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 boundaries.



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

B.4 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 54. (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 55.
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 van Kampen diagram with 0 or 1 hole. Proposition 54 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 any one of its connected components.) 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 56.
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 van Kampen 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 van Kampen 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, in such a manner 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 van Kampen diagram included in D, hence e-narrow by Proposition 55, 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 assumption, 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 57.
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 van Kampen 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 55. 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** (1998), No. 4, 3783–3792.

[AC] G.N. Arzhantseva, P.-A. Cherix, *On the Cayley graph of a generic finitely presented group*, to appear in Bull. Belg. Math. Soc.

[AO] G.N. Arzhantseva, A.Yu. Ol'shanskiĭ, *Generality of the class of groups in which subgroups with a lesser number of generators are free*, Mat. Zametki **59** (1996), No. 4, 489–496; translation in Math. Notes **59** (1996), No. 3–4, 350–355.

[B] K.S. Brown, *Cohomology of groups*, Graduate texts in Mathematics **87**, Springer (1982).

[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** (1995), No. 2, 197–262.

[Ch2] C. Champetier, *Cocroissance des groupes à petite simplification*, Bull. London Math. Soc. **25** (1993), No. 5, 438–444.

[Ch3] C. Champetier, *L'espace des groupes de type fini*, Topology **39** (2000), No. 4, 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*, Geom. Funct. Anal. **13** (2003), No. 1, 73–146.

[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.

[KS] I. Kapovich, P. Schupp, *Genericity, the Arzhantseva-Ol'shanskiĭ method and the isomorphism problem for one-relator groups*, preprint, `arXiv:math.GR/0210307`.

[KSS] I. Kapovich, P. Schupp, V. Shpilrain, *Generic properties of Whitehead's algorithm, stabilizers in Aut*(F_k) *and one-relator groups*, preprint, `arXiv:math.GR/0303386`.

[LS] R.C. Lyndon, P.E. Schupp, *Combinatorial Group Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete **89**, Springer (1977).

[Ols1] A.Yu. Ol'shanskiĭ, *Almost Every Group is Hyperbolic*, Int. J. Algebra Comput. **2** (1992), No. 1, 1–17.

[Ols2] A.Yu. Ol'shanskiĭ, *On residualing homomorphisms and G-subgroups of hyperbolic groups*, Int. J. Algebra Comput. **3** (1993), No. 4, 365–409.

[Oll1] Y. Ollivier, *Critical densities for random quotients of hyperbolic groups*, C.R. Math. Acad. Sci. Paris **336** (2003), No. 5, 391–394.

[Oll2] Y. Ollivier, *Cogrowth and spectral gap of generic groups*, preprint, arXiv:math.GR/0401048

[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] H. 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*, Geom. Funct. Anal. **13** (2003), No. 3, 643–670.

Habilitation à diriger des recherches

Effondrement de quotients aléatoires de groupes hyperboliques avec torsion

Les exemples donnés dans ce texte ont été trouvés à l'extrême fin de ma thèse, et publiés peu après dans une note parue aux Comptes Rendus – Mathématique *de l'Académie des sciences de Paris, volume 341, n°3 (2005), pp. 137-140.*

Yann Ollivier, École normale supérieure de Lyon, 2009

Effondrement de quotients aléatoires de groupes hyperboliques avec torsion

Yann Ollivier

Abstract

We show that random quotients of hyperbolic groups with “harmful” torsion collapse at densities smaller than expected.

Résumé

Nous montrons que les quotients aléatoires de groupes hyperboliques à torsion « meurtrière » s’effondrent à des densités plus petites que prévu.

1 Résultats

Dans un groupe hyperbolique, « adding “sufficiently random” relations to a non-elementary word hyperbolic group gives us a word hyperbolic group again » (M. Gromov, [Gro87], 5.5F). Cette intuition peut être formalisée dans un contexte déterministe (petite simplification relative : [Ch94, Del96]) ou aléatoire. Cette dernière option, retenue dans [Oll04], a l’avantage de permettre de quantifier très précisément le nombre de relations que l’on peut ajouter avant d’obtenir un groupe trivial.

Rappelons le modèle de quotient par des mots aléatoires à densité d. (On renvoie à [Gro93, Ghy03, Oll] pour une discussion générale des groupes aléatoires.) Soit G_0 un groupe hyperbolique non élémentaire, engendré par des éléments $a_1, \ldots, a_m$, $m \geqslant 2$. Pour $0 \leqslant d \leqslant 1$, un *ensemble de mots aléatoires de longueur ℓ à densité d* est l’ensemble aléatoire R obtenu en tirant $(2m)^{d\ell}$ fois de suite un mot au hasard parmi les $(2m)^{\ell}$ mots de longueur ℓ en les $a_i^{\pm 1}$ (on peut aussi considérer seulement les mots réduits, voir [Oll04]). Un *quotient aléatoire de G_0 à densité d et longueur ℓ* est le groupe $G = G_0/\langle R\rangle$ ainsi obtenu.

On dit qu’une propriété de G survient *très probablement* si sa probabilité tend vers 1 lorsque $\ell \to \infty$, à d fixé. Le théorème suivant est extrait de [Oll04] (théorème 4) :

Théorème 1.
Soit G_0 un groupe hyperbolique non élémentaire, à torsion inoffensive (voir ci-dessous), engendré par les éléments $a_1, \ldots, a_m$. Soit $(2m)^{-1/2} < \lambda(G_0) < 1$ le rayon spectral de l’opérateur de marche aléatoire simple sur G_0 engendré par $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$, comme défini dans [Kes59]. Soit $d_{\mathrm{crit}} := -\log_{2m} \lambda(G_0) \in]0; 1/2[$. Alors :

– *si $d < d_{\mathrm{crit}}$, très probablement un quotient aléatoire de G_0 est hyperbolique non-élémentaire ;*
– *si $d > d_{\mathrm{crit}}$, très probablement un quotient aléatoire de G_0 est le groupe trivial $\{e\}$.*

La définition suivante est introduite et discutée dans [Oll04]. Elle est comparable à, mais moins exigeante que, la propriété de « centralisateurs cycliques » utilisée dans [Ch00].

Définition 2.
Un groupe hyperbolique G est dit à torsion inoffensive *si pour tout élément de torsion, son centralisateur est, ou cyclique, ou bien virtuellement $\mathbb{Z}$, ou encore égal à G.*

La raison de cette note est de montrer la nécessité de l'hypothèse de torsion inoffensive :

Théorème 3.
Le théorème 1 ne s'applique pas au groupe hyperbolique $G_0 = (F_4 \times \mathbb{Z}/2\mathbb{Z}) \star F_4$ (pris avec ses neuf générateurs naturels), où $\star$ désigne le produit libre. Plus précisément, il existe une densité $0 < d_{\mathrm{crit}} < -\log_{18} \lambda(G_0)$, telle qu'en densité $d > d_{\mathrm{crit}}$, les quotients aléatoires de G_0 sont très probablement triviaux.

Nous donnons dans la discussion ci-après plus de détails sur le comportement des quotients aléatoires de G_0.

2 Démonstration

L'idée est qu'à partir d'une certaine densité dépendant de la taille du sous-groupe $F_4 \times \mathbb{Z}/2\mathbb{Z}$, le facteur $\mathbb{Z}/2\mathbb{Z}$ deviendra central dans le quotient aléatoire. Au-dessus de cette densité, les quotients aléatoires se comporteront donc comme des quotients de $(F_4 \star F_4) \times \mathbb{Z}/2\mathbb{Z}$, groupe qui a une densité critique plus faible que $(F_4 \times \mathbb{Z}/2\mathbb{Z}) \star F_4$.

Fixons deux entiers strictement positifs n et p ; soient les deux groupes

$$G_1 := (F_n \times \mathbb{Z}/2\mathbb{Z}) \star F_p \quad \text{et} \quad G_2 := (F_n \star F_p) \times \mathbb{Z}/2\mathbb{Z}$$

et choisissons une famille génératrice (notée par abus de la même façon dans ces deux groupes), à savoir $a_1, \ldots, a_n, b_1, \ldots, b_p, u$, où bien sûr les a_i sont les générateurs standard du facteur F_n, les b_i les générateurs standard de F_p, et u engendre le facteur $\mathbb{Z}/2\mathbb{Z}$.

On vérifie aisément que les groupes G_1 et G_2 sont hyperboliques, non élémentaires.

Le rayon spectral du groupe libre F_k est, d'après [Kes59], égal à $\sqrt{2k-1}/k$. Le lemme 4.1 de [Kes59] donne le rayon spectral d'un produit direct, et on obtient $\lambda(F_k \times \mathbb{Z}/2\mathbb{Z}) = (1 + k\lambda(F_k))/(1+k) = \left(1 + \sqrt{2k-1}\right)/(k+1)$. En conséquence de quoi

$$\lambda(G_2) = \left(1 + \sqrt{2n+2p-1}\right)/(n+p+1)$$

et lorsque $n = p = 4$, on obtient $d_2 := -\log_{2(n+p+1)} \lambda(G_2) \approx 0,212$.

On montre maintenant que $\lambda(G_1) < \lambda(G_2)$. L'inégalité large découle bien sûr du fait que G_2 est un quotient de G_1. D'après le théorème 1 de [Kes59], le rayon spectral augmente strictement lors d'un quotient d'un groupe par (la clôture normale d') un sous-groupe non moyennable. Le noyau de l'application quotient $G_1 \to G_2$ contient par construction les deux commutateurs $ub_1u^{-1}b_1^{-1}$ et $ub_2u^{-1}b_2^{-1}$ (si $p \geqslant 2$). Il est facile de vérifier que ces derniers engendrent un sous-groupe libre de rang 2 dans G_1, sous-groupe qui est donc non moyennable d'où l'affirmation.

Nous allons montrer que pour $n = p = 4$, les quotients aléatoires de G_1 sont très probablement triviaux dès que $d > d_2$. Comme $d_1 := -\log_{2(n+p+1)} \lambda(G_1) > d_2$, le théorème 1 contredirait ce fait.

Soit R un ensemble de $18^{d\ell}$ mots de longueur ℓ en $u, a_1, \dots, a_4, b_1, \dots, b_4$ et leurs inverses, choisis au hasard parmi les 18^ℓ possibles, suivant la définition d'un quotient aléatoire à densité d. Passons à l'étude du quotient $G_1/\langle R\rangle$.

Soit C le sous-groupe de G_1 engendré par $u, a_1, \dots, a_4$, et évaluons la probabilité qu'il existe un $r \in R$ qui appartienne à C. Le nombre de mots de longueur ℓ appartenant à C est au moins 10^ℓ (tous les mots en $u, a_1, \dots, a_4$ et leurs inverses). La probabilité qu'un mot aléatoire appartienne à C est donc minorée par $(10/18)^\ell$. Par conséquent si le cardinal de R est beaucoup plus grand que $(18/10)^\ell$, très probablement l'un des éléments de R appartiendra à C. Par définition du modèle, ceci se produit lorsque $18^{d\ell} \gg (18/10)^\ell$ pour $\ell \to \infty$, soit dès que $d > d_C := 1 - \log_{18} 10 \approx 0,203$. À noter que $d_C < d_2$. Notons aussi pour plus tard que $d_C > 0$ car $G_1/\langle C\rangle = F_p$ n'est pas moyennable (critère de [Kes59]).

On sait donc que si $d > d_C$ (ce que l'on suppose désormais), très probablement l'ensemble R contient un élément de C. Le même argument de décompte montre que, pour chaque x dans l'ensemble (fini !) $\{u, a_1, \dots, a_4, b_1, \dots, b_4\}$, l'ensemble R contient très probablement un mot de la forme xc où c est un mot de longueur $\ell - 1$ appartenant à C.

Soit $H = G_1/\langle R\rangle$ le quotient aléatoire à étudier. Soit $r = xc \in R$ avec $c \in C$. Par définition de C, dans G_1 les éléments u et c commutent, et donc dans H on a

$$uxu^{-1}x^{-1} =_{G_1} uxcu^{-1}c^{-1}x^{-1} =_{G_1} uru^{-1}r^{-1} =_H e$$

car, dans H, on a $r = e$ par définition.

Comme, pour $d > d_C$, un tel r existe pour tous les générateurs x de G_1, on en déduit que u commute avec tous les générateurs de H et est donc central dans H. Soit ainsi $S \subset G_1$ l'ensemble de ces commutateurs $\{[u, x]\,;\, x \in \{u, a_1, \dots, a_4, b_1, \dots, b_4\}\}$, on a

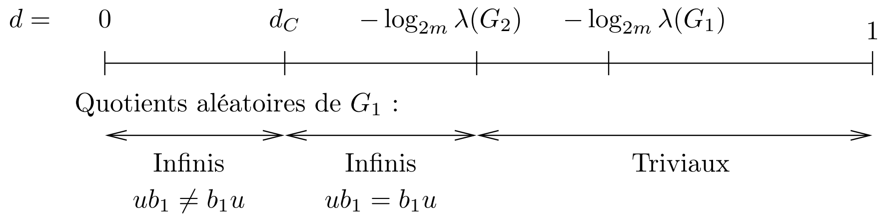
$$H = G_1/\langle R\rangle = G_1/\langle R \cup S\rangle = (G_1/\langle S\rangle)/\langle R\rangle = G_2/\langle R\rangle$$

Mais $G_2/\langle R\rangle$ est un quotient aléatoire de G_2 (en effet la notion de quotient par des mots aléatoires ne dépend pas du groupe considéré mais seulement d'un ensemble de symboles formant les mots). Le groupe G_2 étant à torsion inoffensive (le centralisateur de u est G_2 tout entier), le théorème 1 s'y applique, et donc ses quotients aléatoires sont triviaux dès que $d > d_2$.

Par conséquent, dès que $d > \max(d_C, d_2) = d_2$, et non seulement pour $d > d_1$, les quotients aléatoires de G_1 sont triviaux, ce qui était à démontrer.

3 Discussion

Résumons le comportement des quotients aléatoires de G_1. Pour $0 \leqslant d < d_C \approx 0,203$, on peut montrer que les axiomes de [Oll04] sont satisfaits et que donc les quotients aléatoires se comportent comme décrit dans [Oll04]. Mais pour $d > d_C$, l'axiome 4 de [Oll04] est mis en défaut et les quotients aléatoires de G_1 se comportent comme ceux de G_2, et sont donc triviaux pour $d > d_2 \approx 0,212$ au lieu d'une valeur plus grande attendue. (L'écart entre $0,203$ et $0,212$ peut être augmenté en choisissant de plus grands n et p.)



Les deux phases $d < d_C$ et $d > d_C$ sont réellement différentes : dans la seconde, la relation $ub_1 = b_1 u$ a lieu comme dans G_2, alors qu'elle est fausse dans la première (en effet dans cette phase les axiomes de [Oll04] tiennent et donc le rayon d'injectivité du quotient tend vers l'infini avec ℓ) ; il y a donc une différence observable dans la boule de rayon 1 du graphe de Cayley.

On peut sans aucun doute arranger plus de trois phases en utilisant des groupes tels que

$$(((F_m \times \mathbb{Z}/2\mathbb{Z}) \star F_p) \times \mathbb{Z}/2\mathbb{Z}) \star F_q$$

avec plusieurs densités critiques correspondant aux densités des centralisateurs des différents éléments de torsion.

Il serait intéressant de disposer d'un critère général explicitant dans quels cas la torsion, de non inoffensive, devient « meurtrière » au sens où elle modifie la densité critique. La variable pertinente est sans doute l'exposant asymptotique avec lequel la marche aléatoire voit le centralisateur des éléments de torsion : si cet exposant est supérieur à l'exposant avec lequel elle voit l'élément neutre (qui est $\log \lambda$), alors le quotient aléatoire sera sans doute trop tôt trivial. De même, une théorie similaire utilisant des exposants de croissance plutôt que de cocroissance permettrait d'obtenir des contre-exemples dans le cadre du modèle dit géodésique de quotient aléatoire (théorème 3 de [Oll04]).

Remerciements

Je tiens à remercier Étienne Ghys qui relut et transmit cette note, ainsi que, tout particulièrement, A. Ol'shanskiĭ qui me signala une erreur dans une version préliminaire de [Oll04] où l'hypothèse de torsion inoffensive n'apparaissait pas.

Références

[Ch94] Ch. Champetier, Petite simplification dans les groupes hyperboliques, Ann. Fac. Sci. Toulouse Math., ser. 6, vol. 3 (1994), n° 2, 161–221.

[Ch00] Ch. Champetier, L'espace des groupes de type fini, Topology 39 (2000), n° 4, 657–680.

[Del96] T. Delzant, Sous-groupes distingués et quotients des groupes hyperboliques, Duke Math. J. 83 (1996), n° 3, 661–682.

[Ghy03] É. Ghys, Groupes aléatoires, séminaire Bourbaki 916 (2003).

[Gro87] M. Gromov, Hyperbolic groups, in Essays in group theory, ed. S. M. Gersten, Springer (1987), 75–265.

[Gro93] M. Gromov, Asymptotic invariants of infinite groups, in Geometric group theory, eds. G. Niblo, M. Roller, Cambridge University Press, Cambridge (1993).

[Kes59] H. Kesten, Symmetric random walks on groups, Trans. AMer. Math. Soc. 92 (1959), 336–354.

[Oll04] Y. Ollivier, Sharp phase transition theorems for hyperbolicity of random groups, GAFA, Geom. Funct. Anal. 14 (2004), n° 3, 595–679.

[Oll] Y. Ollivier, A January 2005 invitation to random groups, expository manuscript.

Cogrowth and spectral gap of generic groups

Ce texte a été commencé en 2003 durant ma thèse, et achevé plus tard. Il a été publié dans le volume 55, n°1 (2005) des Annales *de l'Institut Fourier (Grenoble), pp. 289–317.*

Yann Ollivier, École normale supérieure de Lyon, 2009

Cogrowth and spectral gap of generic groups

Yann Ollivier

Abstract

We prove that that for all ε, having cogrowth exponent at most $1/2+\varepsilon$ (in base $2m-1$ with m the number of generators) is a generic property of groups in the density model of random groups. This generalizes a theorem of Grigorchuk and Champetier. More generally we show that the cogrowth of a random quotient of a torsion-free hyperbolic group stays close to that of this group.

This proves in particular that the spectral gap of a generic group is as large as it can be.

Cogrowth of generic groups. The spectral gap of an infinite group (with respect to a given set of generators) is a quantity controlling the speed of convergence of the simple random walk on the group (see [K]); up to parity problems it is equal to the first eigenvalue of the discrete Laplacian. By a formula of Grigorchuk (Theorem 4.1 of [Gri], see also section 1.1 below) this quantity can also be expressed combinatorially by a quantity called *cogrowth*: the smaller the cogrowth, the larger the spectral gap (see also [C]). So this is an important quantity from the combinatorial, probabilistic and operator-algebraic point of view (see [GdlH] or [W] and the references therein for an overview).

In [Gri] (Theorem 7.1) and [Ch93], Grigorchuk and Champetier show that groups defined by a presentation satisfying the small cancellation condition, or a weaker assumption in the case of Champetier, with long enough relators (depending on the number of relators in in the presentation), has a cogrowth exponent arbitrarily close to $1/2$ (the smallest possible value), hence a spectral gap almost as large as that of the free group with same number of generators.

We get the same conclusion for *generic* groups in a precise probabilistic meaning: that of the density model of random groups introduced in [Gro93], which we briefly recall in section 1.2. (Note that in the density model of random groups, if $d>0$ the number of relators is exponentially large and so Grigorchuk's and Champetier's results do not apply). Recall from [Gro93] that above density $d_{\text{crit}}=1/2$, random groups are very probably trivial.

Theorem 1.
Let $0 \leqslant d < 1/2$ be a density parameter and let G be a random group on $m \geqslant 2$ generators at density d and length ℓ.

Then, for any $\varepsilon>0$, the probability that the cogrowth exponent of G lies in the interval $[1/2; 1/2+\varepsilon]$ tends to 1 as $\ell \to \infty$.

In particular, this provides a new large class of groups having a large spectral gap.

This theorem cannot be interpreted by saying that as the relators are very long, the geometry of the group is trivial up to scale ℓ. Indeed, cogrowth is an asymptotic invariant and thus takes into account the very non-trivial geometry of random groups at scale ℓ (see paragraph "locality of cogrowth" below). This is crudely examplified by the collapse of the group when density is too large.

Our primary motivation is the study of generic properties of groups. The study of random groups emerged from an affirmation of Gromov in [Gro87] that "almost every group is hyperbolic". Since the pioneer work of Champetier ([Ch95]) and Ol'shanskiĭ ([Ols]) it has been flourishing, now having connections with lots of topics in group theory such as property T, the Baum-Connes conjecture, small cancellation, the isomorphism problem, the Haagerup property, planarity of Cayley graphs...

The density model of random groups (which we recall in section 1.2), introduced in [Gro93], is very rich in allowing a precise control of the number of relators to be put in the group (and it actually allows this number to be very large). It has proven to be very fruitful, as random groups at different densities can have different properties (e.g. property T). See [Gh] and [Oll] for a general discussion of random groups and the density model, and [Gro93] for an enlightening presentation of the initial intuition behind this model.

Cogrowth of random quotients. A generic group is simply a random quotient of a free group[1]. More generally, we show that, when taking a random quotient of a torsion-free hyperbolic group, the cogrowth of the resulting group is very close to that of the initial group. Recall from [Oll] that a random quotient of a torsion-free hyperbolic group is very probably trivial above some critical density d_{crit}, which precisely depends on the cogrowth of the group (see Theorem 7 in section 1.2 below).

Theorem 2.
Let G_0 be a non-elementary, torsion-free hyperbolic group generated by the elements $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$. Let η be the cogrowth exponent of G_0 with respect to this generating set.

Let $0 \leqslant d < d_{crit}$ be a density parameter and let G be a random quotient (either by plain or reduced random words) of G_0 at density d and length ℓ.

Then, for any $\varepsilon > 0$, the probability that the cogrowth exponent of G lies in the interval $[\eta; \eta + \varepsilon]$ tends to 1 when $\ell \to \infty$.

Of course Theorem 1 is a particular case of Theorem 2. Also, since the cogrowth and gross cogrowth exponent can be computed from each other by the Grigorchuk formula (see section 1.1), this implies that the gross cogrowth exponent does not change either.

[1] There is a very interesting and intriguing parallel approach to generic groups, developed by Champetier in [Ch00], which consists in considering the topological space of all group presentations with a given number of generators. See [P] for a description of connections of this approach with other problems in group theory.

This answers a very natural question arising from [Oll]: indeed, it is known that for each torsion-free hyperbolic group, the critical density d_{crit}, below which random quotients are infinite and above which they are trivial, is equal to 1 minus the cogrowth exponent (resp. 1 minus the gross cogrowth exponent) for a quotient by random reduced words (resp. random plain words). So wondering what happens to the cogrowth exponent after a random quotient is very natural.

Knowing that cogrowth does not change much allows in particular to iterate the operation of taking a random quotient. These iterated quotients are the main ingredient in the construction by Gromov ([Gro03]) of a counter-example to the Baum-Connes conjecture with coefficients (see also [HLS]). Without the stability of cogrowth, in order to get the crucial cogrowth control necessary to build these iterated quotients Gromov had to use a very indirect and non-trivial way involving property T (which allows uniform control of cogrowth over all infinite quotients of a group); this could be avoided with our argument. So besides their interest as generic properties of groups, the results presented here could be helpful in the field.

Remark 3.
Theorem 2 only uses the two following facts: that the random quotient axioms of [Oll] are satisfied, and that there is a local-to-global principle for cogrowth in the random quotient. So in particular the result holds under slightly weaker conditions than torsion-freeness of G_0, as described in [Oll] ("harmless torsion").

Locality of cogrowth in hyperbolic groups. As one of our tools we use a result about locality of cogrowth in hyperbolic groups. Cogrowth is an asymptotic invariant, and large relations in a group can change it noticeably. But in hyperbolic groups, if the hyperbolicity constant is known, it is only necessary to evaluate cogrowth in some ball in the group to get a bound for cogrowth of the group (see Proposition 8). So in this case cogrowth is accessible to computation.

In the case of random quotients by relators of length ℓ, this principle shows that it is necessary to check cogrowth up to words of length at most $A\ell$ for some constant A (which depends on density and actually tends to infinity when d is close to the critical density), so that geometry of the quotient matters up to scale ℓ but not at higher scales.

This result may have independent interest.

About the proofs. The proofs make heavy use of the techniques developed in [Ch93] and [Oll]. We hope to have included precise enough reminders.

As often in hyperbolic group theory, the general case is very involved but lots of ideas are already present in the case of the free group. So in order to help understand the structure of the argument, we first present a proof in the case of the free group (Theorem 1), and then the proof of Theorem 2 for any torsion-free hyperbolic group.

Also, the proofs for random quotients by reduced and plain random words are very similar. They can be treated at once using the general but heavy terminology of [Oll].

We rather chose to present the proof of Theorem 1 in the case of reduced words (for which it seems to be more natural) and of Theorem 2 in the case of plain words.

Acknowledgments. I would like to thank Étienne Ghys and Pierre Pansu for very helpful discussions and many comments on the text. Pierre Pansu especially insisted that I should go on with this question at a time when I had no ideas about it. Lots of the ideas presented here emerged during my stay at the École normale supérieure de Lyon in Spring 2003, at the invitation of Damien Gaboriau and Étienne Ghys. I am very grateful to all the team of the math department there for their warmth at receiving me.

1 Definitions and notations

1.1 Cogrowth, gross cogrowth, spectral gap

These are variants around the same ideas. The spectral radius of the random walk operator on a group was studied by Kesten in [K], and cogrowth was defined later, simultaneously by Grigorchuk ([Gri]) and Cohen ([C]). See [GdlH] for an overview of results and open problems about these quantities and other, related ones.

So let G be an infinite group generated by the elements $a_1^{\pm}, \ldots, a_m^{\pm 1}$. Let W_ℓ be the set of words w of length ℓ in the letters $a_1^{\pm}, \ldots, a_m^{\pm 1}$ such that w is equal to e in the group G. Let $W'_\ell \subset W_\ell$ be the set of *reduced* words in W_ℓ. (Note that W'_ℓ is empty if G is freely generated by $a_1, \ldots, a_m$.) Denote the cardinal of a set by $|.|$.

Definition 4 (Cogrowth exponent).
The cogrowth exponent *of G with respect to $a_1, \ldots, a_m$ is defined as*

$$\eta = \lim_{\substack{\ell \to \infty \\ \ell \text{ even}}} \frac{1}{\ell} \log_{2m-1} \left| W'_\ell \right|$$

or $\eta = 1/2$ if G is freely generated by $a_1, \ldots, a_m$.

The gross cogrowth exponent *of G with respect to $a_1, \ldots, a_m$ is defined as*

$$\theta = \lim_{\substack{\ell \to \infty \\ \ell \text{ even}}} \frac{1}{\ell} \log_{2m} |W_\ell|$$

So the cogrowth exponent is the logarithm in base $2m-1$ of the cogrowth as defined by Grigochuk and Cohen. The exponents η and θ always lie in the interval $[1/2; 1]$, with equality only in case of η of a free group. Amenability of G is equivalent to $\eta = 1$ and to $\theta = 1$.

It is shown in the references mentioned above that the limit exists. We have to take ℓ even in case there are no relations of odd length in the group (in which case W_ℓ is empty).

The convention for the free group is justified by the following Grigorchuk formula ([Gri], Theorem 4.1):

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

which allows to compute one exponent knowing the other (also using that these are at least 1/2), and shows that η and θ vary the same way. Given that θ is well-defined for a free group, the formula yields $\eta(F_m) = 1/2$. As this is also the convention which makes all our statements valid without isolating the case of a free group, we strongly plead for this being the right convention.

The cogrowth exponent is also the exponent of growth of the kernel of the natural map from the free group F_m to G sending a_i to a_i.

The probability of return to e in time t of the simple random walk on G (with respect to the generators $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$) is of course equal to $|W_t|/(2m)^t$. So $(2m)^{\theta-1}$ is also the spectral radius of the random walk operator on $L^2(G)$ defined by $Mf(x) = \frac{1}{2m}\sum f(xa_i^{\pm 1})$. This is the form studied by Kesten ([K]), who denotes by λ this spectral radius.

Since the discrete Laplacian on G is equal to the operator $\mathrm{Id} - M$, $1-(2m)^{\theta-1}$ is also equal to $\min(\lambda_1, 2-\lambda_\infty)$ where λ_1 is the smallest and λ_∞ the largest eigenvalue of the Laplacian acting on $L^2(G)$. (The problems of λ_∞ and of parity of ℓ in the definition can be avoided by considering lazy random walks.) In particular, if θ (or η) is small then the spectral gap λ_1 is large.

The cardinals of the sets W_ℓ of course satisfy the superadditivity property $|W_{\ell+\ell'}| \geqslant |W_\ell|\,|W_{\ell'}|$. This implies that for any ℓ we have an exact (instead of asymptotic) bound $|W_\ell| \leqslant (2m)^{\theta\ell}$. For cogrowth this is not exactly but almost true, due to reduction problems, and we have $\left|W'_{\ell+\ell'+2}\right| \geqslant |W'_\ell|\left|W'_{\ell'}\right|$ and the exact inequality $|W'_\ell| \leqslant (2m-1)^{\eta\ell+2}$. We will often implicitly use these inequalities in the sequel.

1.2 The density model of random groups

A random group is a quotient of a free group $F_m = \langle a_1, \ldots, a_m\rangle$ by (the normal closure of) a randomly chosen set $R \subset F_m$. Typically R is viewed as a set of words in the letters $a_i^{\pm 1}$. So defining a random group is giving a law for R.

More generally, given a group G_0 generated by the elements $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$, and given a set R of random words in these generators we define a random quotient of G_0 by $G = G_0/\langle R\rangle$.

The density model which we now define allows a precise control on the size of R: the bigger the size of R, the smaller the random group. For comparison, remember the number of words of length ℓ in $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$ is $(2m)^\ell$, and the number of reduced words is $(2m)(2m-1)^{\ell-1} \approx (2m-1)^\ell$.

In the whole text we suppose $m \geqslant 2$.

Definition 5 (Density model of random groups or quotients). *Let G_0 be a group generated by the elements $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$. Let $0 \leqslant d \leqslant 1$ be a density parameter.*

Let R be a set of $(2m)^{d\ell}$ randomly chosen words of length ℓ (resp. a set of $(2m-1)^{d\ell}$ randomly chosen reduced words of length ℓ), uniformly and independently picked among all those words.

We call the group $G = G_0/\langle R\rangle$ a random quotient *of G_0 by plain random words (resp. by reduced random words), at density d, at length ℓ.*

In case G_0 is the free group F_m and reduced words are taken, we simply call G a random group.

In this definition, we can also replace "words of length ℓ " by "words of length between ℓ and $\ell + C$ " for any constant C; the theorems presented thereafter remain valid. In [Oll], section 4, we describe generalizations of these models.

The interest of the density model was established by the following theorem of Gromov, which shows a sharp phase transition between infinity and triviality of random groups.

Theorem 6 (M. Gromov, [Gro93]).
Let $d < 1/2$. Then with probability tending to 1 as ℓ tends to infinity, random groups at density d are infinite hyperbolic.

Let $d > 1/2$. Then with probability tending to 1 as ℓ tends to infinity, random groups at density d are either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$.

(The occurrence of $\mathbb{Z}/2\mathbb{Z}$ is of course due to the case when ℓ is even; this disappears if one takes words of length between ℓ and $\ell + C$ with $C \geqslant 1$.)

Basically, $d\ell$ is to be interpreted as the "dimension" of the random set R (see the discussion in [Gro93]). As an illustration, if $L < 2d\ell$ then very probably there will be two relators in R sharing a common subword of length L. Indeed, the dimension of the couples of relators in R is $2d\ell$, whereas sharing a common subword of length L amounts to L "equations", so the dimension of those couples sharing a subword is $2d\ell - L$, which is positive if $L < 2d\ell$. This "shows" in particular that at density d, the small cancellation condition $C'(2d)$ is satisfied.

Since a random quotient of a free group is hyperbolic, one can wonder if a random quotient of a hyperbolic group is still hyperbolic. The answer is basically yes, and the critical density in this case is linked to the cogrowth exponent of the initial group.

Theorem 7 (Y. Ollivier, [Oll]).
Let G_0 be a non-elementary, torsion-free hyperbolic group, generated by the elements $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$, with cogrowth exponent η and gross cogrowth exponent θ.

Let $0 \leqslant d \leqslant 1$ be a density parameter, and set $d_{crit} = 1 - \theta$ (resp. $d_{crit} = 1 - \eta$).

If $d < d_{crit}$, then a random quotient of G_0 by plain (resp. reduced) random words is infinite hyperbolic, with probability tending to 1 as ℓ tends to infinity.

If $d > d_{crit}$, then a random quotient of G_0 by plain (resp. reduced) random words is either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$, with probability tending to 1 as ℓ tends to infinity.

This is the context in which Theorem 2 is to be understood.

1.3 Hyperbolic groups and isoperimetry of van Kampen diagrams

Let G be a group given by the finite presentation $\langle\, a_1, \dots, a_m \mid R\,\rangle$. Let w be a word in the $a_i^{\pm 1}$'s. We denote by $|w|$ the number of letters of w, and by $\|w\|$ the distance from e to w in the Cayley graph of the presentation, that is, the minimal length of a word representing the same element of G as w.

Let λ be the maximal length of a relation in R.

We refer to [LS] for the definition and basic properties of van Kampen diagrams. Remember that a word represents the neutral element of G if and only if it is the boundary word of some van Kampen diagram. If D is a van Kampen diagram, we denote its number of faces by $|D|$ and its boundary length by $|\partial D|$.

It is known ([Sh]) that G is hyperbolic if and only if there exists a constant $C_1 > 0$ such that for any (reduced) word w representing the neutral element of G, there exists a van Kampen diagram with boundary word w, and with at most $|w|/C_1$ faces. This can be reformulated as: for any word w representing the neutral element of G, there exists a van Kampen diagram with boundary word w satisfying the isoperimetric inequality

$$|\partial D| \geqslant C_1 \, |D|$$

We are going to use a *homogeneous* way to write this inequality. The above form compares the boundary length of a van Kampen diagram to its number of faces. This amounts to comparing a length with a number, which is not very well-suited for geometric arguments, especially when dealing with groups having relations of very different lengths.

So let D be a van Kampen diagram w.r.t. the presentation and define the *area* of D to be

$$\mathcal{A}(D) = \sum_{f \text{ face of } D} |\partial f|$$

which is also the number of external edges (not couting “filaments”) plus twice the number of internal ones. This has, heuristically speaking, the homogeneity of a length.

It is immediate to see that if D satisfies $|\partial D| \geqslant C_1 \, |D|$, then we have $|\partial D| \geqslant C_1 \, \mathcal{A}(D)/\lambda$ (recall λ is the maximal length of a relation in the presentation). Conversely, if $|\partial D| \geqslant C_2 \, \mathcal{A}(D)$, then $|\partial D| \geqslant C_2 \, |D|$. So we can express the isoperimetric inequality using $\mathcal{A}(D)$ instead of $|D|$.

Say a diagram is *minimal* if it has minimal area for a given boundary word. So G is hyperbolic if and only if there exists a constant $C > 0$ such that every minimal van Kampen diagram satisfies the isoperimetric inequality

$$|\partial D| \geqslant C \, \mathcal{A}(D)$$

This formulation is homogeneous, that is, it compares a length to a length. This inequality is the one that naturally arises in $C'(\alpha)$ small cancellation theory (with $C = 1 - 6\alpha$) as well as in random groups at density d (with $C = \frac{1}{2} - d$). So in these contexts the value of C is naturally linked with some parameters of the presentation.

This kind of isoperimetric inequality is also the one appearing in the assumptions of Champetier in [Ch93], in random quotients of hyperbolic groups (cf. [Oll]) and in the (infinitely presented) limit groups constructed by Gromov in [Gro03]. So we think this is the right way to write the isoperimetric inequality when the lengths of the relators are very different.

2 Locality of cogrowth in hyperbolic groups

The goal of this section is to show that in a hyperbolic group, in order to estimate cogrowth (which is an asymptotic invariant), it is enough to check only words of bounded length, where the bound depends on the quality of the isoperimetric inequality in the group.

Everything here is valid, *mutatis mutandis*, for cogrowth and gross cogrowth.

Here $G = \langle a_1, \ldots, a_m \mid R \rangle$ ($m \geqslant 2$) is a hyperbolic group and W_ℓ is the set of reduced words of length ℓ in the $a_i^{\pm 1}$ equal to e in G. Let also λ be the maximal length of a relation in R.

As explained above, hyperbolicity of G amounts to the existence of some constant $C > 0$ such that any minimal van Kampen diagram D over this presentation satisfies the isoperimetric inequality

$$|\partial D| \geqslant C\mathcal{A}(D)$$

We will prove the following.

Proposition 8.
Suppose that, for some $A > 1$, for any $A\lambda/4 \leqslant \ell \leqslant A\lambda$ one has

$$|W_\ell| \leqslant (2m-1)^{\eta\ell}$$

for some $\eta \geqslant 1/2$.

Then for any $\ell \geqslant A\lambda/4$,

$$|W_\ell| \leqslant (2m-1)^{\eta\ell(1+o(1)_{A\to\infty})}$$

where the constant implied in $o(1)$ depends only on C.

It follows from the proof that actually $1 + o(1) \leqslant \exp \frac{200}{C\sqrt{A}}$, so that is it enough to take $A \approx 40000/C^2$ for a good result.

Proof.
First we need some simple lemmas.

The *distance to boundary* of a face of a van Kampen diagram is the minimal length of a sequence of faces adjacent by an edge, beginning with the given face and ending with a face adjacent to the boundary (so that a boundary face is at distance 1 from the boundary).

Set $\alpha = 1/\log(1/(1-C)) \leqslant 1/C$, where we can suppose $C \leqslant 1$.

Lemma 9.
Let D be a minimal van Kampen diagram. Then D can be written as a disjoint union

$D = D_1 \cup D_2$ (with maybe D_2 not connected) such that each face of D_1 is at distance at most $\alpha \log(\mathcal{A}(D)/\lambda)$ from the boundary of D, and D_2 has area at most λ.

Proof.
Since D is minimal it satisfies the isoperimetric inequality $|\partial D| \geqslant C\mathcal{A}(D)$. Thus, the cumulated area of the faces of D which are adjacent to the boundary is at least $C\mathcal{A}(D)$, and so the cumulated area of the faces at distance at least 2 from the boundary is at most $(1-C)\mathcal{A}(D)$.

Applying the same reasoning to the (maybe not connected) diagram obtained from D by removing the boundary faces, we get by induction that the cumulated area of the faces of D lying at distance at least k from the boundary is at most $(1-C)^{k-1}\mathcal{A}(D)$. Taking $k = 1+\alpha \log(\mathcal{A}(D)/\lambda)$ (rounded up to the nearest integer) provides the desired decomposition. □

In the sequel we will neglect divisibility problems (such as the length of a diagram being a multiple of 4).

Lemma 10.
Let D be a minimal van Kampen diagram. D can be partitioned into two diagrams D', D'' by cutting it along a path of length at most $\lambda + 2\alpha\lambda \log(\mathcal{A}(D)/\lambda)$ such that each of D' and D'' contains at least one quarter of the boundary of D.

(Here a *path* in a diagram is meant to be a path in its 1-skeleton.)

Proof.
Consider the decomposition $D = D_1 \cup D_2$ of the previous lemma, and first suppose that D_2 is empty, so that any face of D_1 lies at distance at most $\alpha\lambda \log(\mathcal{A}(D)/\lambda)$ from the boundary.

Let L be the boundary length of D and mark four points A, B, C, D on ∂D at distance $L/4$ of each other. As D is $\alpha \log(\mathcal{A}(D)/\lambda)$-narrow, there exists a path of length at most $2\alpha\lambda \log(\mathcal{A}(D)/\lambda)$ joining either a point of AB to a point of CD or a point of AD to a point of BC, which provides the desired cutting.

Now if D_2 was not empty, first retract each connected component of D_2 to a point: the reasoning above shows that there exists a path of length at most $2\alpha\lambda \log(\mathcal{A}(D)/\lambda)$ joining either a point of AB to a point of CD or a point of AD to a point of BC, *not counting the length in D_2*. But since the sum of the lengths of the faces of D_2 is at most λ, the cumulated length of the travel in D_2 is at most λ, hence the lemma. □

The cardinal of the W_ℓ's (almost in the case of cogrowth, see above) satisfy the supermultiplicativity property $|W_\ell| \geqslant |W_{\ell-L}|\,|W_L|$. Using narrowness of diagrams we are able to show a converse inequality, which will enable us to control cogrowth.

Corollary 11.
We have, up to parity problems,

$$\begin{aligned} |W_\ell| &\leqslant \sum_{\ell/4 \leqslant \ell' \leqslant 3\ell/4} \left|W_{\ell'+2\alpha\lambda\log(\ell/C\lambda)+\lambda}\right| \left|W_{\ell-\ell'+2\alpha\lambda\log(\ell/C\lambda)+\lambda}\right| \\ &\leqslant \frac{\ell}{\lambda} \max_{\ell/4 \leqslant \ell' \leqslant 3\ell/4} \left|W_{\ell'+2\alpha\lambda\log(\ell/C\lambda)+3\lambda}\right| \left|W_{\ell-\ell'+2\alpha\lambda\log(\ell/C\lambda)+3\lambda}\right| \end{aligned}$$

Proof.
Any word in W_ℓ is the boundary word of some (minimal) van Kampen diagram D with boundary length ℓ, and so the first inequality follows from the previous lemma, together with the inequality $\mathcal{A}(D) \leqslant |\partial D|/C$.

The last inequality uses the fact that, up to moving the cutting points by at most λ, we can assume that the lengths involved are multiples of λ, hence the factor ℓ/λ in front of the max and the increase of the lengths by 2λ. □

Now for the proof of Proposition 8 proper.

First, choose ℓ between $A\lambda$ and $4A\lambda/3$. By Corollary 11 and the assumptions, we have

$$|W_\ell| \leqslant (2m-1)^{\eta(\ell+4\alpha\lambda\log(\ell/C\lambda)+6\lambda)+\log_{2m-1}(\ell/\lambda)}$$

Let B be a number (depending on C) such that

$$4\alpha\log(B/C) + 6 + \frac{1}{\eta}\log_{2m-1} B \leqslant B$$

(noting that $m \geqslant 2$, $\eta \geqslant 1/2$ and $\alpha \leqslant 1/C$ one can check that $B \geqslant 144/C^2$ is enough). It is then easy to check that for $B' \geqslant B$ one has

$$4\alpha\log(B'/C) + 6 + \frac{1}{\eta}\log_{2m-1} B' \leqslant 2\sqrt{B'B}$$

Thus, if $\ell \geqslant A\lambda$ and $A \geqslant B$ we have

$$|W_\ell| \leqslant (2m-1)^{\eta\left(\ell+2\lambda\sqrt{AB}\right)} \leqslant (2m-1)^{\eta\ell\left(1+2\sqrt{B/A}\right)}$$

We have just shown that if $|W_\ell| \leqslant (2m-1)^{\eta\ell}$ for $\ell \leqslant A\lambda$, then $|W_\ell| \leqslant (2m-1)^{\eta\ell\left(1+2\sqrt{B/A}\right)}$ for $\ell \leqslant (4A/3)\lambda$. Thus, iterating the process shows that for $\ell \leqslant (4/3)^k A\lambda$ we have

$$|W_\ell| \leqslant (2m-1)^{\eta\ell\prod_{0\leqslant i<k}\left(1+2\sqrt{\frac{B}{A}}\left(\frac{3}{4}\right)^{i/2}\right)}$$

and we are done as the product $\prod_i\left(1+2\sqrt{\frac{B}{A}}\left(\frac{3}{4}\right)^{i/2}\right)$ converges to some value tending to 1 when $A \to \infty$; if one cares, its value is less than $\exp\frac{200}{C\sqrt{A}}$. □

3 Application to random groups: the free case

Here we first treat the case when the initial group G_0 is the free group F_m on m generators. This will serve as a template for the more complex general case.

So let $G = \langle a_1, \ldots, a_m \mid R\rangle$ be a random group at density d, with R a set of $(2m-1)^{d\ell}$ random reduced words.

We have to evaluate the number of reduced words of a given length L which represent the trivial element in G. Any such word is the boundary word of some van Kampen diagram D with respect to the set of relators R. We will proceed as follows: for any diagram D involving n relators, we will evaluate the expected number of n-tuples of random relators from R that make it a van Kampen diagram. We will show that this expected number is controlled by the boundary length L of the diagram, and this will finally allow to control the number of van Kampen diagrams of boundary length L.

We call a van Kampen diagram *non-filamenteous* if each of its edges lies on the boundary on some face. Each diagram can be decomposed into non-filamenteous components linked by filaments. For the filamenteous part we will use the estimation from [Ch93], one step of which counts the number of ways in which the different non-filamenteous parts can be glued together to form a van Kampen diagram.

So we first focus on non-filamenteous diagrams, for which a genuinely new argument has to be given compared to [Ch93] (since the number of relators here is unbounded).

We first suppose that we care only about diagrams with at most K faces, for some K to be chosen later. (We will of course use the locality of cogrowth principle to remove this assumption.)

3.1 Fulfilling of diagrams

So let D be a non-filamenteous van Kampen diagram. Let $|D|$ be its number of faces and let $n \leqslant |D|$ be the number of different relators it involves. Let m_i, $1 \leqslant i \leqslant n$ be the number of times the i-th relator appears in D, where we choose to enumerate the relators in decreasing order of multiplicity, that is, $m_1 \geqslant m_2 \geqslant \ldots \geqslant m_n$. Let also D_i be the subdiagram of D made of relators $1, 2, \ldots, i$ only, so that $D = D_n$.

It is shown in [Oll] (section 2.2) that to this diagram we can associate numbers $d_1, \ldots, d_n$ such that

- The probability that i given random relators fulfill D_i is less than $(2m-1)^{d_i - id\ell}$; consequently, the probability that there exists an i-tuple of relators in R fulfilling D_i is less than $(2m-1)^{d_i}$.
- The following isoperimetric inequality holds :

$$|\partial D| \geqslant (1-2d)\ell\,|D| + 2\sum d_i(m_i - m_{i+1})$$

So for a given n-tuple of random relators, the probability that this n-tuple fulfills D is at most $(2m-1)^{\inf(d_i - id\ell)}$. So, as there are $(2m-1)^{nd\ell}$ n-tuples of relators in R, the expected number S of n-tuples fulfilling D in R is at most

$$S \leqslant (2m-1)^{nd\ell + \inf(d_i - id\ell)}$$

which so turns out to be not only an upper bound for the probability of D to be fulfillable but rather an estimate of the number of ways in which it is. (The probabilities that two n-tuples fulfill the diagram are independent only when the n-tuples are disjoint, but expectation is linear anyway.)

Set $d'_i = d_i - id\ell$. Then, rewriting the isoperimetric inequality above and using that $m_i - m_{i+1} \geqslant 0$ yields

$$\begin{aligned}
|\partial D| &\geqslant (1-2d)\ell\,|D| + 2\sum(d'_i + id\ell)(m_i - m_{i+1}) \\
&= (1-2d)\ell\,|D| + 2d\ell\sum m_i + 2\sum d'_i(m_i - m_{i+1}) \\
&= \ell\,|D| + 2(\inf d'_i)\sum(m_i - m_{i+1}) + 2\sum(d'_i - \inf d'_i)(m_i - m_{i+1}) \\
&\geqslant \ell\,|D| + 2m_1 \inf d'_i \\
&\geqslant \ell\,|D| + 2\inf d'_i
\end{aligned}$$

and consequently

$$\mathbb{E}S \leqslant (2m-1)^{nd\ell + \inf d'_i} \leqslant (2m-1)^{|D|d\ell + \frac{1}{2}(|\partial D| - |D|\ell)} = (2m-1)^{\frac{1}{2}(|\partial D| - (1-2d)\ell|D|)}$$

Of course this also holds for filamenteous diagrams because the faces are the same but $|\partial D|$ is even greater. So the conclusion is:

Proposition 12.
For any reduced van Kampen diagram D, the expected number of ways it can be fulfilled by random relators at density d is at most $(2m-1)^{\frac{1}{2}(|\partial D| - (1-2d)\ell|D|)}$.

By Markov's inequality, the probability to pick a random presentation R for which $S \geqslant (2m-1)^{\varepsilon\ell}\mathbb{E}S$ is less than $(2m-1)^{-\varepsilon\ell}$. Since the number of diagrams with less than K faces grows subexponentially in ℓ, we have shown:

Proposition 13.
For any fixed integer K and any $\varepsilon > 0$, with probability exponentially close to 1 as $\ell \to \infty$, for each (non-filamenteous) van Kampen diagram with at most K faces, the number of ways to fulfill it with relators of R is at most $(2m-1)^{\frac{1}{2}(|\partial D| - (1-2d-\varepsilon)\ell|D|)}$.

In particular, taking $\varepsilon < (\frac{1}{2} - d)/2$, this is less than $(2m-1)^{|\partial D|/2}$.

3.2 Evaluation of the cogrowth

We now conclude using the general scheme of [Ch93], together with Proposition 8 which allows to check only a finite number of diagrams.

Consider a reduced word w in the generators $a_i^{\pm 1}$, representing e in the random group. This word is the boundary word of some van Kampen diagram D which may have filaments.

Choose $\varepsilon > 0$. We are going to show that with probability exponentially close to 1 when $\ell \to \infty$, the number of such words w is at most $(2m-1)^{(1/2+\varepsilon)|w|}$.

We know from [Oll] (Section 2.2) that up to exponentially small probability in ℓ, we can suppose that any diagram satisfies the inequality

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

where C depends only on the density d (basically $C = 1/2 - d$ divided by the constants appearing in the Cartan-Hadamard-Gromov theorem, see [Oll]) and not on ℓ.

Now we use Proposition 8. We are facing a group G in which all relations are of length ℓ. Consider a constant A given by Proposition 8 such that if we know that $|W_L| \leqslant (2m-1)^{L(1/2+\varepsilon/2)}$ for $L \leqslant A\ell$, then we know that $|W_L| \leqslant (2m-1)^{L(1/2+\varepsilon)}$ for any L. Such an A depends only on the isoperimetry constant C.

So we suppose that our word w has length at most $A\ell$. We have $|w| = |\partial D| \geqslant C\ell\,|D|$ and in particular, $|D| \leqslant A/C$, which is to say, we have to consider only diagrams with a number of faces bounded independently of ℓ.

So set $K = A/C$, which most importantly does not depend on ℓ. After Proposition 13, we can assume (up to exponentially small probability) that for any non-filamenteous diagram D' with at most K faces, the number of ways to fulfill it with relators of the random presentation is at most $(2m-1)^{|\partial D'|/2}$.

Back to our word w read on the boundary of some diagram D. Decompose D into filaments and connected non-filamenteous parts D_i. The word w is determined by the following data: a set of relators from the random presentation R fulfilling the D_i's, a set of reduced words to put on the filaments, the combinatorial choice of the diagrams D_i, and the combinatorial choice of how to connect the D_i's using the filaments.

The combinatorial part is precisely the one analyzed in [Ch93]. It is shown there (section "Premier pas") that if each D_i satisfies $|\partial D_i| \geqslant L$, the combinatorial factor controlling the connecting of the D_i's by the filaments and the sharing of the length $|\partial D|$ between the filaments and the D_i's is less than

$$\frac{|w|}{L}\,|w|\,(eL)^{2|w|/L}(2eL)^{|w|/L}(3eL)^{2|w|/L}$$

Observe that for L large enough this behaves like $(2m-1)^{|w|O(\log L/L)}$.

Here each diagram D_i satisfies $|\partial D_i| \geqslant C\ell\,|D_i| \geqslant C\ell$, so setting $L = C\ell$, each D_i has boundary length at least L. In particular, $O(\log L/L) = O(\log \ell/\ell)$.

The number of components D_i is obviously at most $|w|/L$. Each component has at most K faces since D itself has. So the number of choices for the combinatorial choices of the diagrams D_i's is at most $N(K)^{|w|/L}$ where $N(K)$ is the (finite!) number of planar graphs with at most K faces. This behaves like $(2m-1)^{|w|O(1/L)}$.

Now the number of ways to fill the D_i's with relators from the random presentation is, after Proposition 13, at most $\prod (2m-1)^{|\partial D_i|/2} = (2m-1)^{\sum |\partial D_i|/2}$.

The last choice to take into account is the choice of reduced words to put on the filaments. The total length of the filaments is $\frac{1}{2}(|w| - \sum |\partial D_i|)$ (each edge of a filament counts twice in the boundary), thus the number of ways to fill in the filaments is at most $(2m-1)^{\frac{1}{2}(|w| - \sum |\partial D_i|)}$.

So the total number of possibilities for w is

$$(2m-1)^{|w|O(\log \ell/\ell) + \frac{1}{2}(|w| - \sum |\partial D_i|) + \sum |\partial D_i|/2}$$

and if we take ℓ large enough, this will be at most $(2m-1)^{|w|(1/2+\varepsilon/2)}$, after what we conclude by Proposition 8.

This proves Theorem 1.

4 The non-free case

Now we deal with random quotients of a non-elementary torsion-free hyperbolic group G_0. We are going to give the proof in the case of a random quotient by plain random words, the case of a quotient by random reduced words being similar.

So let G_0 be a non-elementary torsion-free hyperbolic group given by the presentation $\langle\, a_1, \ldots, a_m \mid Q \,\rangle$ ($m \geqslant 2$), with the relations in Q having length at most λ. Let θ be the gross cogrowth of G_0 w.r.t. this generating set. Let $G = G_0/\langle R\rangle$ be a random quotient of G_0 by a set R of $(2m)^{d\ell}$ randomly chosen words of length ℓ. Also set $\beta = 1 - \theta$, so that the random quotient axioms of [Oll] (section 4) are satisfied.

We have to show that the number of boundary words of van Kampen diagrams of a given boundary length L grows slower than $(2m)^{L(\theta+\varepsilon)}$. This time, since we are going to give a proof in the case of gross cogrowth rather than cogrowth, we will not have many problems with filaments: the counting of filaments is already included in the knowledge of gross cogrowth of G_0.

For a van Kampen diagram D, let D'' be the subdiagram made of faces bearing "new" relators in R, and D' be the part made of faces bearing "old" relators in Q. By Proposition 32 of [Oll], we know that very probably G is hyperbolic and that its isoperimetric inequality takes the form

$$|\partial D| \geqslant \kappa\ell \left|D''\right| + \kappa' \left|D'\right|$$

whenever D is reduced and D' is minimal, with $\kappa, \kappa' > 0$ and where, most importantly, κ and κ' do not depend on ℓ. By definition of $\mathcal{A}(D)$, this can be rewritten as $|\partial D| \geqslant C\mathcal{A}(D)$ with $C = \min(\kappa, \kappa'/\lambda)$.

Fix some $\varepsilon > 0$ and let A be the constant provided by Proposition 8 applied to G, having the property that if we know that gross cogrowth is at most $\theta + \varepsilon/2$ up to words of length $A\ell$, then we know that gross cogrowth is at most $\theta + \varepsilon$. This A depends on ε, C and G_0 but not on ℓ. Thanks to this and the isoperimetric inequality, we only have to consider diagrams of boundary length at most $A\ell$ hence area at most $A\ell/C$. In particular the number of new relators $|D''|$ is at most A/C. So for all the sequel set

$$K = A/C$$

which, most importantly, does not depend on ℓ. This is the maximal size of diagrams we have to consider, thanks to the local-global principle.

4.1 Reminder from [Oll]

In this context, it is proven in [Oll] that the van Kampen diagram D can be seen as a "van Kampen diagram at scale ℓ with respect to the new relators, with equalities modulo G_0". More precisely, this can be stated as follows: (we refer to [Oll] for the definition of "strongly reduced" diagrams; the only thing to know here is that for any word equal to e in G, there exists a strongly reduced van Kampen diagram with this word as its boundary word).

Proposition 14 ([Oll], section 6.6).
Let $G_0 = \langle\, S \mid Q \,\rangle$ be a non-elementary hyperbolic group, let R be a set of words of length ℓ, and consider the group $G = G_0/\langle R\rangle = \langle\, S \mid Q \cup R \,\rangle$.

Let $K \geqslant 1$ be an arbitrarily large integer and let $\varepsilon_1, \varepsilon_2 > 0$ be arbitrarily small numbers. Take ℓ large enough depending on $G_0, K, \varepsilon_1, \varepsilon_2$.

Let D be a van Kampen diagram with respect to the presentation $\langle\, S \mid Q \cup R \,\rangle$, which is strongly reduced, of area at most $K\ell$. Let also D' be the subdiagram of D which is the union of the 1-skeleton of D and of those faces of D bearing relators in Q (so D' is a possibly non-simply connected van Kampen diagram with respect to G_0), and suppose that D' is minimal.

We will call worth-considering *such a van Kampen diagram.*

Let $w_1, \ldots, w_p$ be the boundary (cyclic) words of D', so that each w_i is either the boundary word of D or a relator in R.

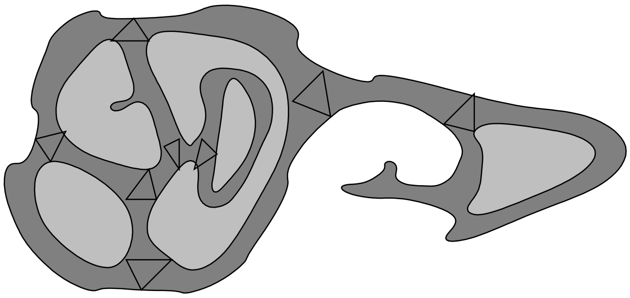
Then there exists an integer $k \leqslant 3K/\varepsilon_2$ and words $x_2, \ldots, x_{2k+1}$ such that:

- *Each x_i is a subword of some cyclic word w_j;*
- *As subwords of the w_j's, the x_i's are disjoint and their union exhausts a proportion at least $1 - \varepsilon_1$ of the total length of the w_j's.*
- *For each $i \leqslant k$, there exists words δ_1, δ_2 of length at most $\varepsilon_2(|x_{2i}| + |x_{2i+1}|)$ such that $x_{2i}\delta_1 x_{2i+1}\delta_2 = e$ in G_0.*
- *If two words x_{2i}, x_{2i+1} are subwords of the boundary words of two faces of D bearing the same relator $r^{\pm 1} \in R$, then, as subwords of r, x_{2i} and x_{2i+1} are either disjoint or equal with opposite orientations (so that the above equality reads $x\delta_1 x^{-1}\delta_2 = e$).*

The couples (x_{2i}, x_{2i+1}) are called translators. *Translators are called* internal, internal-boundary *or* boundary-boundary *according to whether x_{2i} and x_{2i+1} is a subword of some w_j which is a relator in R or the boundary word of D.*

(There are slight differences between the presentation here and that in [Oll]. Therein, boundary-boundary translators did not have to be considered: they were eliminated earlier in the process, before section 6.6, because they have a positive contribution to boundary length, hence always improve isoperimetry and do not deserve consideration in order to prove hyperbolicity. Moreover, in [Oll] we further distinguished "commutation translators" for the kind of internal translator with $x_{2i} = x_{2i+1}^{-1}$, which we need not do here.)

Translators appear as dark strips on the following figure:



Remark 15.
The number of ways to partition the words w_i into translators is at most $(2K\ell)^{12K/\varepsilon_2}$, because each w_i can be determined by its starting- and endpoint, which can be given as numbers between 1 and $2K\ell$ which is an upper bound for the cumulated length of the w_i's (since the area of D is at most $K\ell$). For fixed K and ε_2 this grows subexponentially in ℓ.

Remark 16.
Knowing the words x_i, the number of possibilities for the boundary word of the diagram is at most $(6K/\varepsilon_2)!$ (choose which subwords x_i make the boundary word of the diagram, in which order), which does not depend on ℓ for fixed K and ε_2.

We need another notion from [Oll], namely, that of *apparent length* of an element in G_0. This basically answers the question: If this element were obtained through a random walk at time t, what would be a reasonable value of t? This accounts for the fact that, unlike in the free group, the hitting probability of an element in the group does not depend only on the norm of this element.

Apparent length is defined in [Oll] in a more general setting, with respect to a measure on the group, which is here the measure obtained after a simple random walk with respect to the given set of generators $a_1, \ldots, a_m$. We only give here what the definition amounts to in our context.

Definition 17 (Definition 36 of [Oll]).
Let x be a word. Let $\varepsilon_2 > 0$. Let L be an integer. Let $p_L(xuyv = e)$ be the probability that, for a random word y of length L, there exists elements $u, v \in G_0$ of norm at most $\varepsilon_2(|x| + L)$ such that $xuyv = e$ in G_0.

The apparent length *of x at test-length L is*

$$\mathbb{L}_L(x) = -\frac{1}{1-\theta} \log_{2m} p_L(xuyv = e) - L$$

The apparent length *of x is*

$$\mathbb{L}(x) = \min \left(\|x\| \frac{\theta}{1-\theta}, \min_{0 \leqslant L \leqslant K\ell} \mathbb{L}_L(x) \right)$$

where we recall ℓ is the length of the relators in a random presentation.

(The first term $\|x\|\,\theta/(1-\theta)$ is an easy upper bound for $\mathbb{L}_{\|x\|}(x)$, and so if $\|x\| \leqslant K\ell$ then the first term in the min is useless.)

It is shown in [Oll], section 6.7, that in a randomly chosen presentation at density d and length ℓ, all subwords of the relators have apparent length at most 4ℓ, with probability exponentially close to 1 as $\ell \to \infty$. So from now on we suppose that this is indeed the case.

We further need the notion of a *decorated abstract van Kampen diagram* (which was implicitly present in the free case when we mentioned the probability that some diagram "is fulfilled by random relators"), which is inspired by Proposition 14: it carries the combinatorial information about how the relators and boundary word of a diagram were cut into subwords in order to make the translators.

Definition 18 (Decorated abstract van Kampen diagram). *Let $K \geqslant 1$ be an arbitrarily large integer and let $\varepsilon_1, \varepsilon_2 > 0$ be arbitrarily small numbers. Let I_ℓ be the cyclically ordered set of ℓ elements.*

A decorated abstract van Kampen diagram $\mathcal{D}$ *is the following data:*

- *An integer $|\mathcal{D}| \leqslant K$ called its* number of faces.
- *An integer $|\partial\mathcal{D}| \leqslant K\ell$ called its* boundary length.
- *An integer $n \leqslant |\mathcal{D}|$ called its* number of distinct relators.
- *An application $r^{\mathcal{D}}$ from $\{1, \ldots, |\mathcal{D}|\}$ to $\{1, \ldots, n\}$; if $r^{\mathcal{D}}(i) = r^{\mathcal{D}}(j)$ we will say that* faces i and j bear the same relator.
- *An integer $k \leqslant 3K/\varepsilon_2$ called the* number of translators *of $\mathcal{D}$.*
- *For each integer $2 \leqslant i \leqslant 2k+1$, a set of the form $\{j_i\} \times I'_i$ where either j_i is an integer between 1 and $|\mathcal{D}|$ and I'_i is an oriented cyclic subinterval of I_ℓ, or $j_i = |\mathcal{D}| + 1$ and I'_i is a subinterval of $I_{|\partial\mathcal{D}|}$; this is called an* (internal) subword of the j_i-th face *in the first case, or a* boundary subword *in the second case.*
- *For each integer $1 \leqslant i \leqslant k$ such that $j_{2i} \leqslant |\mathcal{D}|$, an integer between 0 and 4ℓ called the* apparent length of the $2i$-th subword.

such that

- *The sets $\{j_i\} \times I'_i$ are all disjoint and the cardinal of their union is at least $(1-\varepsilon_1)\,(|\mathcal{D}|\,\ell + |\partial\mathcal{D}|)$.*
- *For all $1 \leqslant i \leqslant k$ we have $j_{2i} \leqslant j_{2i+1}$ (this can be ensured by maybe swapping them).*
- *If two faces j_{2i} and j_{2i+1} bear the same relator, then either I'_{2i} and I'_{2i+1} are disjoint or are equal with opposite orientations.*

This way, Proposition 14 ensures that any worth-considering van Kampen diagram D with respect to $G_0/\langle R\rangle$ defines a decorated abstract van Kampen diagram $\mathcal{D}$ in the way suggested by terminology (up to rounding the apparent lengths to the nearest integer; we neglect this problem). We will say that $\mathcal{D}$ is *associated to* D. Remark 15 tells that the number of decorated abstract van Kampen diagrams grows subexponentially with ℓ (for fixed K).

Given a decorated abstract van Kampen diagram $\mathcal{D}$ and n given relators $r_1, \ldots, r_n$, we say that these relators *fulfill* $\mathcal{D}$ if there exists a worth-considering van Kampen diagram D with respect to $G_0/\langle r_1, \ldots, r_n\rangle$, such that the associated decorated abstract van Kampen diagram is $\mathcal{D}$. Intuitively speaking, the relators $r_1, \ldots, r_n$ can be "glued modulo G_0 in the way described by $\mathcal{D}$".

So we want to study which diagrams can probably be fulfilled by random relators in R. The main conclusion from [Oll] is that these are those with large boundary length, hence hyperbolicity of the quotient $G_0/\langle R\rangle$. Here for cogrowth we are rather interested in the number of ways to fulfill an abstract diagram with given boundary length.

4.2 Cogrowth of random quotients

So now let R again be a set of $(2m)^{d\ell}$ random relators. Let $\mathcal{D}$ be a given decorated abstract van Kampen diagram. Recall we set $K = A/C$. The free parameters ε_1 and ε_2 will be chosen later.

We will show (Proposition 21) that, up to exponentially small probability in ℓ, the number of different boundary words of worth-considering van Kampen diagrams D such that $\mathcal{D}$ is associated to D, is at most $(2m)^{\theta|\partial\mathcal{D}|(1+\varepsilon/2)}$.

Further notations. Let n be the number of distinct relators in $\mathcal{D}$. For $1 \leqslant a \leqslant n$, let m_a be the number of times the a-th relator appears in $\mathcal{D}$. Up to reordering, we can suppose that the m_a 's are non-increasing. Also to avoid trivialities take n minimal so that $m_n \geqslant 1$.

Let also P_a be the probability that, if a words $r_1, \ldots, r_a$ of length ℓ are picked at random, there exist $n - a$ words $r_{a+1}, \ldots, r_n$ of length ℓ such that the relators $r_1, \ldots, r_n$ fulfill $\mathcal{D}$. The P_a 's are of course a non-increasing sequence of probabilities. In particular, P_n is the probability that a random n-tuple of relators fulfills $\mathcal{D}$.

Back to our set R of $(2m)^{d\ell}$ randomly chosen relators. Let P^a be the probability that there exist a relators $r_1, \ldots, r_a$ in R, such that there exist words $r_{a+1}, \ldots, r_n$ of length ℓ such that the relators $r_1, \ldots, r_n$ fulfill $\mathcal{D}$. Again the P^a 's are a non-increasing sequence of probabilities and of course we have

$$P^a \leqslant (2m)^{ad\ell} P_a$$

since the $(2m)^{ad\ell}$ factor accounts for the choice of the a-tuple of relators in R.

The probability that there exists a worth-considering van Kampen diagram D with respect to the random presentation R, such that $\mathcal{D}$ is associated to D, is by definition less than P^a for any a. In particular, if for some $\mathcal{D}$ we have $P^a \leqslant (2m)^{-\varepsilon'\ell}$, then with probability exponentially close to 1 when $\ell \to \infty$, $\mathcal{D}$ is not associated to any worth-considering van Kampen diagram of the random presentation. Since, by Remark 15, the number of possibilities for $\mathcal{D}$ grows subexponentially with ℓ, we can sum this over $\mathcal{D}$ and conclude that for any $\varepsilon' > 0$, with probability exponentially close to 1 when $\ell \to \infty$ (depending on ε'), all decorated abstract van Kampen diagrams $\mathcal{D}$ associated to some worth-considering van Kampen diagram of the random presentation satisfy $P^a \geqslant (2m)^{-\varepsilon'\ell}$ and in particular

$$P_a \geqslant (2m)^{-ad\ell-\varepsilon'\ell}$$

which we assume from now on.

We need to define one further quantity. Keep the notations of Definition 18. Let $1 \leqslant a \leqslant n$ and let $1 \leqslant i \leqslant k$ where k is the number of translators of $\mathcal{D}$. Say that the i-th translator is half finished at time a if $r^{\mathcal{D}}(j_{2i}) \leqslant a$ and $r^{\mathcal{D}}(j_{2i+1}) > a$, that is, if one side of the translator is a subword of a relator $r_{a'}$ with $a' \leqslant a$ and the other of $r_{a''}$ with $a'' > a$. Now let A_a be the sum of the apparent lengths of all translators which are half finished at time a. In particular, A_n is the sum of the apparent lengths of all subwords $2i$ such that $2i$ is an internal subword and $2i+1$ is a boundary subword of $\mathcal{D}$.

The proof. In this context, equation ($\star$) (section 6.8) of [Oll] reads

$$A_a - A_{a-1} \geqslant m_a \left(\ell(1-\varepsilon'') + \frac{\log_{2m} P_a - \log_{2m} P_{a-1}}{\beta} \right)$$

where ε'' tends to 0 when our free parameters $\varepsilon_1, \varepsilon_2$ tend to 0 (and ε'' also absorbs the $o(\ell)$ term in [Oll]). Also recall that in the model of random quotient by plain random words, we have

$$\beta = 1 - \theta$$

by Proposition 15 of [Oll].

Setting $d'_a = \log_{2m} P_a$ and summing over a we get, using $\sum m_a = |\mathcal{D}|$, that

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

Now recall we saw above that for any $\varepsilon' > 0$, taking ℓ large enough we can suppose that $P_a \geqslant (2m)^{-ad\ell-\varepsilon'\ell}$, that is, $d'_a + ad\ell + \varepsilon'\ell \geqslant 0$. Hence

$$\begin{aligned} A_n &\geqslant |\mathcal{D}|\, \ell(1-\varepsilon'') + \frac{1}{\beta} \sum (d'_a + ad\ell + \varepsilon'\ell)(m_a - m_{a+1}) \\ &\quad - \frac{1}{\beta} \sum (ad\ell + \varepsilon'\ell)(m_a - m_{a+1}) \\ &= |\mathcal{D}|\, \ell(1-\varepsilon'') + \frac{1}{\beta} \sum (d'_a + ad\ell + \varepsilon'\ell)(m_a - m_{a+1}) - \frac{d\ell}{\beta} \sum m_a - \frac{\varepsilon'\ell}{\beta} m_1 \\ &\geqslant |\mathcal{D}|\, \ell(1-\varepsilon'') + \frac{d'_n + nd\ell + \varepsilon'\ell}{\beta} m_n - \frac{d\ell + \varepsilon'\ell}{\beta} \sum m_a \end{aligned}$$

where the last inequality follows from the fact that we chose the order of the relators so that $m_a - m_{a+1} \geqslant 0$.

So using $m_n \geqslant 1$ we finally get

$$A_n \geqslant |\mathcal{D}|\,\ell\left(1-\varepsilon''-\frac{d+\varepsilon'}{\beta}\right)+\frac{d'_n+nd\ell}{\beta}$$

Suppose the free parameters ε_1, ε_2 and ε' are chosen small enough so that $1 - \varepsilon'' - (d+\varepsilon')/\beta \geqslant 0$ (remember that ε'' is a function of $\varepsilon_1, \varepsilon_2$ and K; we will further decrease ε_1 and ε_2 later). This is possible since by assumption we take the density d to be less than the critical density β. This is the only, but crucial, place where density plays a role. Thus the first term in the inequality above is non-negative and we obtain the simple inequality $A_n \geqslant (d'_n + nd\ell)/\beta$.

Proposition 19.
Up to exponentially small probability in ℓ, we can suppose that any worth-considering decorated abstract van Kampen diagram $\mathcal{D}$ satisfies

$$A_n(\mathcal{D}) \geqslant \frac{d'_n(\mathcal{D})+nd\ell}{\beta}$$

This we now use to evaluate the number of possible boundary words for van Kampen diagrams associated with $|\mathcal{D}|$.

Remember that, by definition, d'_n is the log-probability that n random relators $r_1, \ldots, r_n$ fulfill $\mathcal{D}$. As there are $(2m)^{nd\ell}$ n-tuples of random relators in R (by definition of the density model), by linearity of expectation the expected number of n-tuples of relators in R fulfilling $\mathcal{D}$ is $(2m)^{nd\ell+d'_n}$, hence the interest of an upper bound for $d'_n + nd\ell$.

By the Markov inequality, for given $\mathcal{D}$ the probability to pick a random set R such that the number of n-tuples of relators of R fulfilling $\mathcal{D}$ is greater than $(2m)^{nd\ell+d'_n+C\varepsilon\ell/4}$, is less than $(2m)^{-C\varepsilon\ell/4}$. By Remark 15 the number of possibilities for $\mathcal{D}$ is subexponential in ℓ, and so, using Proposition 19 we get

Proposition 20.
Up to exponentially small probability in ℓ, we can suppose that for any worth-considering decorated abstract van Kampen diagram $\mathcal{D}$, the number of n-tuples of relators in R fulfilling $\mathcal{D}$ is at most

$$(2m)^{\beta A_n(\mathcal{D})+C\varepsilon\ell/4}$$

Now let D be a van Kampen diagram associated to $\mathcal{D}$. Given $\mathcal{D}$ we want to evaluate the number of different boundary words for D. Recall Proposition 14: the boundary word of D is determined by giving two words for each boundary-boundary translator, and one word for each internal-boundary translator, this latter one being subject to the apparent length condition imposed in the definition of $\mathcal{D}$. By Remark 16, the number of ways to combine these subwords into a boundary word for D is controlled by K and ε_2 (independently of ℓ).

So let (x_{2i}, x_{2i+1}) be a boundary-boundary translator in D. By Proposition 14 (definition of translators) there exist words δ_1, δ_2 of length at most $\varepsilon_2(|x_{2i}| + |x_{2i+1}|)$ such that $x_{2i}\delta_1 x_{2i+1}\delta_2 = e$ in G_0. So $x_{2i}\delta_1 x_{2i+1}\delta_2$ is a word representing the trivial element in G_0, and by definition of θ the number of possibilities for (x_{2i}, x_{2i+1}) is at most $(2m)^{\theta(|x_{2i}|+|x_{2i+1}|)(1+2\varepsilon_2)}$.

Now let (x_{2i}, x_{2i+1}) be an internal-boundary translator. The apparent length of x_{2i} is imposed in the definition of $\mathcal{D}$. The subword x_{2i} is an internal subword of D, and so by definition is a subword of some relator $r_i \in R$. So if the relators in D are given, x_{2i} is determined. But knowing x_{2i} still leaves open lots of possibilities for x_{2i+1}. This is where apparent length comes into play.

Since $y = x_{2i+1}$ is a boundary word of D one has $|y| \leqslant A\ell \leqslant K\ell$. So by definition we have $\mathbb{L}(x_{2i}) \leqslant \mathbb{L}_{|y|}(x_{2i})$. By definition of translators there exist words u and v of length at most $\varepsilon_2\ell$ such that $x_{2i}uyv = e$ in G_0. By definition of $\mathbb{L}_{|y|}(x_{2i})$, if y' is a random word of length $|y|$, then the probability that $x_{2i}uy'v = e$ in G_0 is $(2m)^{-(1-\theta)\left(|y|+\mathbb{L}_{|y|}(x_{2i})\right)} \leqslant (2m)^{-(1-\theta)(|y|+\mathbb{L}(x_{2i}))}$. This means that the total number of words y' of length $|y|$ such that there exists u, v with $x_{2i}uyv = e$ is at most $(2m)^{|y|}(2m)^{-(1-\theta)(|y|+\mathbb{L}(x_{2i}))} = (2m)^{\theta|y|-(1-\theta)\mathbb{L}(x_{2i})}$. So, given x_{2i}, the number of possibilities for $y = x_{2i+1}$ is at most this number.

So if the relators in R fulfilling $\mathcal{D}$ are fixed, the number of possible boundary words for D is the product of $(2m)^{\theta(|x_{2i}|+|x_{2i+1}|)(1+2\varepsilon_2)}$ for all boundary-boundary translators (x_{2i}, x_{2i+1}), times the product of $(2m)^{\theta|x_{2i+1}|-(1-\theta)\mathbb{L}(x_{2i})}$ for all internal-boundary translators (x_{2i}, x_{2i+1}), times the number of ways to order these subwords (which is subexponential in ℓ by Remark 16), times the number of possibilities for the parts of the boundary of D not belonging to any translator, which by Proposition 14 have total length not exceeding $\varepsilon_1 K\ell$.

Now the sum of $|x_{2i}| + |x_{2i+1}|$ for all boundary-boundary translators (x_{2i}, x_{2i+1}), plus the sum of $|x_{2i+1}|$ for all internal-boundary translators, is $|\partial\mathcal{D}|$ (maybe up to $\varepsilon_1 K\ell$). And the sum of $\mathbb{L}(x_{2i})$ for all internal-boundary translators is A_n by definition.

So given $\mathcal{D}$ and given a n-tuple of relators fulfilling $\mathcal{D}$, the number of possibilities for the boundary word of D is at most

$$(2m)^{\theta|\partial\mathcal{D}|(1+2\varepsilon_2)-(1-\theta)A_n+\varepsilon_1 K\ell}$$

up to a subexponential term in ℓ. By Proposition 20 (remember $\beta = 1 - \theta$), if we include the choices of the relators fulfilling $\mathcal{D}$ the number of possibilities is at most

$$(2m)^{\theta|\partial\mathcal{D}|(1+2\varepsilon_2)+\varepsilon_1 K\ell+C\varepsilon\ell/4}$$

If we choose $\varepsilon_2 \leqslant \varepsilon/16$ and $\varepsilon_1 \leqslant \varepsilon C/8K$ so that (using $|\partial D| \geqslant C\ell\,|\mathcal{D}| \geqslant C\ell$ for any fulfillable abstract diagram) the sum of the corresponding terms is less than $\varepsilon\,|\partial D|\,/4$ (note that this choice does not depend on ℓ) and if we remember that, after Remark 15, the number of choices for $\mathcal{D}$ is subexponential in ℓ, we finally get:

Proposition 21.
Up to exponentially small probability in ℓ, the number of different boundary words of worth-considering van Kampen diagrams of a random presentation with given boundary length L, is at most
$$(2m)^{\theta L(1+\varepsilon/2)}$$

But remember the discussion at the beginning of section 4 (where we invoked Proposition 8): it is enough to show that gross cogrowth is at most $\theta+\varepsilon/2$ for words of length L between $A\ell/4$ and $A\ell$. Any such word is the boundary word of a van Kampen diagram of area at most $K\ell$, hence is the boundary word of some worth-considering van Kampen diagram. This ends the proof of Theorem 2.

References

[C] J.M. Cohen, *Cogrowth and Amenability of Discrete Groups*, J. Funct. Anal. **48** (1982), 301–309.

[Ch93] C. Champetier, *Cocroissance des groupes à petite simplification*, Bull. London Math. Soc. **25** (1993), No 5, 438–444.

[Ch95] C. Champetier, *Propriétés statistiques des groupes de présentation finie*, J. Adv. Math. **116** (1995), No 2, 197–262.

[Ch00] C. Champetier, *L'espace des groupes de type fini*, Topology **39** (2000), No 4, 657–680.

[GdlH] R.I. Grigorchuk, P. de la Harpe, *On problems related to growth, entropy, and spectrum in group theory*, Dynam. Control Systems **3** (1997), No 1, 51–89.

[Gh] É. Ghys, *Groupes aléatoires*, séminaire Bourbaki **916** (2003).

[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.

[Gro87] M. Gromov, *Hyperbolic Groups*, in *Essays in group theory*, ed. S.M. Gersten, Springer (1987), 75–265.

[Gro93] M. Gromov, *Asymptotic Invariants of Infinite Groups*, in *Geometric group theory*, ed. G. Niblo, M. Roller, Cambridge University Press, Cambridge (1993).

[Gro03] M. Gromov, *Random Walk in Random Groups*, Geom. Funct. Anal. **13** (2003), No 1, 73–146.

[HLS] N. Higson, V. Lafforgue, G. Skandalis, *Counterexamples to the Baum-Connes conjecture*, Geom. Funct. Anal. **12** (2002), No 2, 330–354.

[K] H. Kesten, *Symmetric Random Walks on Groups*, Trans. Amer. Math. Soc. **92** (1959), 336–354.

[LS] R.C. Lyndon, P.E. Schupp, *Combinatorial Group Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete **89**, Springer (1977).

[Oll] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, GAFA, Geom. Funct. Anal. **14** (2004), No 3, 595–679.

[Ols] A.Yu. Ol'shanskiĭ, *Almost Every Group is Hyperbolic*, Int. J. Algebra Comput. **2** (1992), No 1, 1–17.

[P] F. Paulin, *Sur la théorie élémentaire des groupes libres*, Séminaire Bourbaki **922** (2003).

[Sh] H. Short et al., in *Group Theory from a Geometrical Viewpoint*, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991).

[W] W. Woess, *Random Walks on Infinite Graphs and Groups*, Cambridge Tracts in Mathematics **138**, Cambridge University Press (2000).

Growth exponent of generic groups

Ce texte a été commencé en 2003 durant ma thèse, et achevé plus tard. Il a été publié dans le volume 81, n°3 (2006) des Commentarii Mathematici Helvetici, *pp. 569–593.*

Yann Ollivier, École normale supérieure de Lyon, 2009

Growth exponent of generic groups

Yann Ollivier

Abstract

In [GrH97], Grigorchuk and de la Harpe ask for conditions under which some group presentations have growth rate close to that of the free group with the same number of generators. We prove that this property holds for a generic group (in the density model of random groups). Namely, for every positive ε, the property of having growth exponent at least $1-\varepsilon$ (in base $2m-1$ where m is the number of generators) is generic in this model. In particular this extends a theorem of Shukhov [Shu99].

More generally, we prove that the growth exponent does not change much through a random quotient of a torsion-free hyperbolic group.

Introduction

The growth exponent is a very natural quantity associated to a group presentation, measuring the rate of growth of the balls in the group with respect to some given set of generators. Namely, let $G = \langle\, a_1, \ldots, a_m \mid R \,\rangle$ be a finitely generated group. For $\ell \geqslant 0$ let $B_\ell \subset G$ be the set of elements of norm at most ℓ with respect to this generating set. The *growth exponent* of G (sometimes called *entropy*) with respect to this set of generators is

$$g = \lim_{\ell\to\infty} \frac{1}{\ell} \log_{2m-1} |B_\ell|$$

The maximal value of g is 1, which is achieved if and only if G is the free group F_m on the m generators $a_1, \ldots, a_m$. The limit in the definition exists thanks to the submultiplicativity property $|B_{\ell+\ell'}| \leqslant |B_\ell|\,|B_{\ell'}|$. By standard properties of subadditive (or submultiplicative) sequences, this implies in particular that for any ℓ we have $|B_\ell| \geqslant (2m-1)^{g\ell}$.

Growth exponents of groups, first introduced by Milnor, are related to many other properties, for example in Riemannian geometry, dynamical systems and of course combinatorial group theory. We refer to [GrH97], [Har00] (chapters VI and VII), or [Ver00] for some surveys and applications.

The authors of [GrH97] ask for conditions under which some families of groups (namely one-relators groups) have growth exponents getting arbitrarily close to the maximal value 1. Shukhov gave an example of such a condition in [Shu99]: it is proven therein that if a group presentation has long relators satisfying the $C'(1/6)$

small cancellation condition, and if there are “not too many” relators (in a precise sense), then the growth exponent of the group so presented is arbitrarily close to 1.

We prove that having growth exponent at least $1-\varepsilon$ is a *generic* property in the density model of random groups.

For a general discussion and extensive bibliography on random groups and the various models we refer to [Oll-b] or [Gh03]. The density model was introduced by Gromov in [Gro93]. We recall the precise definition in Section 1.1 below; basically, depending on a density parameter $d \geqslant 0$, it consists in taking a group presentation with m fixed generators and $(2m-1)^{d\ell}$ relators taken at random among all reduced words of length ℓ in the generators, and letting $\ell \to \infty$. The intuition is that at density d, any reduced word of length $d\ell$ will appear as a subword of some relator in the presentation.

This model allows a precise control of the quantity of relations put in the random group, which is examplified by the phase transition theorem proven in [Gro93]: below density 1/2, random groups are very probably infinite and hyperbolic, and very probably trivial above density 1/2 (see Theorem 5 below).

Keeping this in mind, our theorem reads:

Theorem 1.
Let $d < 1/2$ be a density parameter and let G be a random group on $m \geqslant 2$ generators at density d and at length ℓ.

Then, for any $\varepsilon > 0$, the probability that the growth exponent of G is at least $1-\varepsilon$ tends to 1 as $\ell \to \infty$.

When $d < 1/12$ this is a consequence of Shukhov's theorem: indeed for densities at most 1/12, random groups satisfy the $C'(1/6)$ small cancellation condition. But for larger densities they do not any more, and so the theorem really provides a large class of new groups with large growth exponent.

Random groups at length ℓ look like free groups at scales lower than ℓ (more precisely, the length of the shortest relation in a random group is ℓ if $d < 1/4$ and $\ell(2-4d-\varepsilon)$ if $d \geqslant 1/4$), and so the cardinality of balls of course grows with exponent 1 at the beginning. However, growth is an asymptotic invariant, and the geometry of random groups at scale ℓ is highly non-trivial, so the theorem cannot be interpreted by simply saying that random groups look like free groups at small scales.

More generally, we show that for torsion-free hyperbolic groups, the growth exponent is stable in the following sense: if we randomly pick elements in the group and quotient by the normal subgroup they generate (the so-called quotient by *random elements* as opposed to the quotient by randomly picked words in the generators; see details below), then the growth exponent stays almost unchanged, unless we killed too many elements and get the trivial group. Note however that this exponent cannot stay exactly the same, as Arzhantseva and Lysenok proved in [AL02] that quotienting a hyperbolic group by an infinite normal subgroup decreases the growth exponent.

The study of random quotients of hyperbolic groups arises naturally from the knowledge that a random group (a random quotient of the free group) is hyperbolic:

one can wonder whether a random quotient of a hyperbolic group stays hyperbolic. The answer from [Oll04] is yes (see Section 1.1 below for details) up to some critical density equal to $g/2$ where g is the growth exponent of the initial group; above this critical density the random quotient collapses. In this framework our second theorem reads:

Theorem 2.
Let G_0 be a non-elementary torsion-free hyperbolic group of growth exponent g. Let $d < g/2$. Let G be a quotient of G_0 by random elements at density d and at length ℓ.

Then, for any $\varepsilon > 0$, with probability tending to 1 as $\ell \to \infty$, the growth exponent of G lies between $g - \varepsilon$ and g.

Of course, Theorem 1 is just Theorem 2 applied to a free group.

Remark 3.
The proof of Theorem 2 only uses the two following facts: that the random quotient axioms of [Oll04] are satisfied, and that there is a local-to-global principle for growth in the random quotient. So in particular the result holds under slightly weaker conditions than torsion-freeness of G_0, as described in [Oll04] ("harmless torsion").

Locality of growth in hyperbolic groups. As one of our tools we use a result about locality of growth in hyperbolic groups (see the Appendix). Growth is an asymptotic invariant, and large relations in a group can change it noticeably. But in hyperbolic groups, if the hyperbolicity constant is known, it is only necessary to evaluate growth in some ball in the group to get that the growth of the group is not too far from this evaluation (see Proposition 17 in the Appendix).

In the case of random quotients by relators of length ℓ, this principle shows that it is necessary to check growth up to words of length at most $A\ell$ for some large constant A (which depends on density and actually tends to infinity when d is close to the critical density), so that geometry of the quotient matters up to scale ℓ (including the non-trivial geometry of the random quotient at this scale) but not at higher scales.

This result may have independent interest.

About the proofs, and about cogrowth. The proofs presented here make heavy use of the terminology and results from [Oll04]. We have included a reminder (Section 2.2) so that this paper is self-contained.

This paper comes along with a "twin" paper about *co*growth of random groups ([Oll05]). Let us insist that, although the inspiration for these two papers is somewhat the same (use some locality principle and count van Kampen diagrams), the details do differ, except for the reminder from [Oll04] which is identical. Especially, the proof of the locality principle for growth and cogrowth is not at all the same. The counting of van Kampen diagrams begins similarly but soon diverges as we are not evaluating the same things eventually. And we do not work in the same variant of the density model: for growth we use the element variant, whereas for cogrowth we use the word variant (happily these two variants coincide in the case of a free group, that is, for "plain" random groups).

The result of [Oll05] already implies some lower bound for the growth exponent of a random group, thanks to the formula $(2m-1)^{g/2} \geqslant (2m)^{1-\theta}$ where by definition $(2m)^{\theta-1}$ is the spectral radius of the simple random walk operator (see [GrH97]). However this bound is not sharp: for a random group it reads $(2m-1)^g \geqslant m^2/(2m-1) - \varepsilon$ whereas we prove here that $(2m-1)^g \geqslant 2m-1-\varepsilon$.

Acknowledgments. I would like to thank Étienne Ghys, Pierre de la Harpe, Pierre Pansu and an anonymous referee for helpful discussions and many comments on the text. Lots of the ideas presented here emerged during my very nice stay at the École normale supérieure de Lyon in Spring 2003, at the invitation of Damien Gaboriau and Étienne Ghys.

1 Definitions and notations

1.1 Random groups and density

The interest of random groups is twofold: first, to study which properties of groups are *generic*, i.e. shared by a large proportion of groups; second, to provide examples of new groups with given properties. This article falls under both approaches.

A random group is given by a random presentation, that is, the quotient of a free group $F_m = \langle a_1, \ldots, a_m \rangle$ by (the normal closure of) a randomly chosen set $R \subset F_m$. Defining a random group is giving a law for the random set R.

More generally, a random quotient of a group G_0 is the quotient of G_0 by (the normal closure of) a randomly chosen subset $R \subset G_0$.

The philosophy of random groups was introduced by Gromov in [Gro87] through a statement that "almost every group is hyperbolic", the proof of which was later given by Ol'shanskiĭ ([Ols92]) and independently by Champetier ([Ch91, Ch95]). Gromov later defined the density model in [Gro93], in order to precisely control the quantity of relators put in a random group.

Since then random groups have gained broad interest and are connected to lots of topics in geometric or combinatorial group theory (such as the isomorphism problem, property T, Haagerup property, small cancellation, spectral gaps, the Baum-Connes conjecture...), especially since Gromov used them ([Gro03]) to build a counter-example to the Baum-Connes conjecture with coefficients (see also [HLS02]). We refer to [Oll-b] or [Gh03] for a general discussion on random groups and an extensive bibliography.

We now define the density model of random groups. In this model the random set of relations R depends on a density parameter d: the larger d, the larger R. This model exhibits a phase transition between infiniteness and triviality depending on the value of d; moreover, in the infinite phase some properties of the resulting group (such as the rank, property T or the Haagerup property) do differ depending on d, hence the interest of this model.

Definition 4 (Density model of quotient by random elements). *Let G_0 be a group generated by the elements $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$ ($m \geqslant 2$). Let $B_\ell \subset G_0$ be the ball of radius ℓ in G_0 with respect to this generating set.*

Let $d \geqslant 0$ be a density parameter.

Let R be a set of $(2m-1)^{d\ell}$ randomly chosen elements of B_ℓ, uniformly and independently picked in B_ℓ.

We call the group $G = G_0/\langle R\rangle$ a quotient of G_0 by random elements*, at density d and at length ℓ.*

In case G_0 is the free group F_m we simply call G a random group *at density d and at length ℓ.*

We sometimes also refer to this model as the *geodesic model* of random quotients.

In this definition, we can also replace B_ℓ by the sphere S_ℓ of elements of norm exactly ℓ, or by the annulus of elements of norm between ℓ and $\ell + C$ for some constant C: this does not affect our theorems. Compare Theorem 3 in [Oll04].

Another variant (the word variant) of random quotients consists in taking for R a set of reduced (or plain) random words in the generators $a_i^{\pm 1}$, which leads to a different probability distribution. Fortunately in the case of the free group, there is no difference between taking at random elements in B_ℓ or reduced words, so that the notions of random group and of a generic property of groups are well-defined anyway.

Quotienting by elements rather than words seems better suited to control the growth of the quotient (one works with elements of the group all the way long). However, the author believes that the same kind of proof would also work in the word model of random quotients, with a slightly more difficult argument.

The interest of the density model was established by the following theorem of Gromov, which shows a sharp phase transition between infinity and triviality of random groups.

Theorem 5 ([Gro93]).
Let $d < 1/2$. Then with probability tending to 1 as ℓ tends to infinity, random groups at density d are infinite hyperbolic.

Let $d > 1/2$. Then with probability tending to 1 as ℓ tends to infinity, random groups at density d are either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$.

(The occurrence of $\mathbb{Z}/2\mathbb{Z}$ is of course due to the case when ℓ is even and we take elements in the sphere S_ℓ; this disappears if one takes elements in B_ℓ, or of length between ℓ and $\ell + C$ with $C \geqslant 1$.)

Basically, $d\ell$ is to be interpreted as the "dimension" of the random set R (see the discussion in [Gro93]). As an illustration, if $L < 2d\ell$ then very probably there will be two relators in R sharing a common subword of length L. Indeed, the dimension of the pairs of relators in R is $2d\ell$, whereas sharing a common subword of length L amounts to L "equations", so the dimension of those pairs sharing a subword is $2d\ell - L$, which is positive if $L < 2d\ell$. This "shows" in particular that at density d, the small cancellation condition $C'(2d)$ is satisfied.

Since a random quotient of a free group is hyperbolic, one can wonder if a random quotient of a hyperbolic group is still hyperbolic. The answer is basically yes, and for the random elements variant, the critical density is in this case linked to the growth exponent of the initial group.

Theorem 6 ([Oll04], Theorem 3).
Let G_0 be a non-elementary, torsion-free hyperbolic group, generated by the elements $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$, with growth exponent g. Let $0 \leqslant d \leqslant g$ be a density parameter.

If $d < g/2$, then a random quotient of G_0 by random elements at density d is infinite hyperbolic, with probability tending to 1 as ℓ tends to infinity.

If $d > g/2$, then a random quotient of G_0 by random elements at density d is either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$, with probability tending to 1 as ℓ tends to infinity.

This is the context in which Theorem 2 is to be understood.

1.2 Hyperbolic groups and isoperimetry of van Kampen diagrams

Let G be a group given by the finite presentation $\langle a_1, \ldots, a_m \mid R \rangle$. Let w be a word in the $a_i^{\pm 1}$'s. We denote by $|w|$ the number of letters of w, and by $\|w\|$ the distance from e to w in the Cayley graph of the presentation, that is, the minimal length of a word representing the same element of G as w.

Let λ be the maximal length of a relation in R.

We refer to [LS77] for the definition and basic properties of van Kampen diagrams. If D is a van Kampen diagram, we denote its number of faces by $|D|$ and its boundary length by $|\partial D|$.

It is well-known (see for example [Sho91]) that G is hyperbolic if and only if there exists a constant $C_1 > 0$ such that for any word w representing the neutral element of G, there exists a van Kampen diagram with boundary word w satisfying the isoperimetric inequality

$$|\partial D| \geqslant C_1 \, |D|$$

We are going to use a slightly different way to write this inequality. Let D be a van Kampen diagram w.r.t. the presentation and define the *area* of D to be

$$\mathcal{A}(D) = \sum_{f \text{ face of } D} |\partial f|$$

which is also the number of external edges (not couting "filaments") plus twice the number of internal ones. Say a diagram is *minimal* if it has minimal area for a given boundary word.

It is immediate to see that if D satisfies $|\partial D| \geqslant C_1 \, |D|$, then we have $|\partial D| \geqslant C_1 \, \mathcal{A}(D)/\lambda$ (recall λ is the maximal length of a relation in the presentation). Conversely, if $|\partial D| \geqslant C_2 \, \mathcal{A}(D)$, then $|\partial D| \geqslant C_2 \, |D|$. So G is hyperbolic if and only if there exists a constant $C > 0$ such that every minimal van Kampen diagram satisfies the isoperimetric inequality

$$|\partial D| \geqslant C \, \mathcal{A}(D)$$

(where necessarily $C \leqslant 1$ unless G is free).

This inequality naturally arises in $C'(\alpha)$ small cancellation theory (with $C = 1 - 6\alpha$), in random groups at density d (with $C = \frac{1}{2} - d$, see [Oll-a]), in the assumptions of Champetier in [Ch93], in random quotients of hyperbolic groups (cf. [Oll04]) and in the (infinitely presented) limit groups constructed by Gromov in [Gro03]. Moreover there is a nice inequality between C and the hyperbolicity constant δ (Proposition 7 below).

The key feature of this formulation is that both $\mathcal{A}(D)$ and $|\partial D|$ scale the same way when the lengths of the relators change. This homogeneity property is crucial in our applications. So we think this is the right way to write the isoperimetric inequality when the lengths of the relators are very different.

Proposition 7.
Suppose that a hyperbolic group G given by some finite presentation satisfies the isoperimetric inequality

$$|\partial D| \geqslant C\, \mathcal{A}(D)$$

for all minimal van Kampen diagrams D, for some constant $C > 0$.

Let λ be the maximal length of a relation in the presentation. Then the hyperbolicity constant δ of G satisfies

$$\delta \leqslant 12\lambda/C^2$$

Proof.
This is just a careful rewriting of classical proofs. Actually the proof of this is strictly included in [Sho91] (Theorem 2.5). Indeed, what the authors of [Sho91] prove is always of the form "the number of edges in D is at least something, so the number of faces of D is at most this thing divided by ρ" (in their notation ρ is the maximal length of a relation). Reasoning directly with the number of edges instead of the number of faces $|D|$ simplifies their arguments. But $\mathcal{A}(D)$ is simply twice the number of internal edges of D plus the number of boundary edges of D, so it is greater than the number of edges of D.

So simply by removing the seventh sentence in their proof of Lemma 2.6 (where the number of 2-cells of a diagram is evaluated by dividing the number of 1-cells by the maximal length of a relator ρ), we get a new Lemma 2.6 which reads (we stick to their notation in the framework of their proving Theorem 2.5)

Lemma 2.6 of [Sho91].
If $\varepsilon > \rho$, then there is a constant C_1 depending solely on ε, such that the number of 1-cells in $N(\theta)$ is at least $\ell(\theta)\varepsilon/\rho - C_1$. Namely we can set $C_1 = \varepsilon(\varepsilon + \rho)/\rho$.

Similarly, removing the last sentence of their proof of Lemma 2.7 we get a new version of it:

Lemma 2.7 of [Sho91].
If $\varepsilon > \rho$, there is a constant C_2 depending solely on ε such that

$$\mathcal{A}(D) > (\alpha + \beta + \gamma)\varepsilon/\rho - C_2 + 2r$$

where $\mathcal{A}(D)$ is the area of the diagram D. Namely we can set $C_2 = 3C_1 + 4\varepsilon + 2$.

We insist that those modified lemmas are obtained by *removing* some sentences in their proofs, and that there really is nothing to modify.

We still have to re-write the conclusion. In their notation α, β and γ are (up to 4ε) the lengths of the sides of some triangle which, by contradiction, is supposed not to be r-thin (we want to show that if r is large enough, then every triangle is r-thin).

The assumption $|\partial D| \geqslant C\,\mathcal{A}(D)$ reads

$$\mathcal{A}(D) \leqslant (\alpha+\beta+\gamma)/C + 12\varepsilon/C$$

Combining this inequality and the result of Lemma 2.7, we have

$$(\alpha+\beta+\gamma)\varepsilon/\rho - C_2 + 2r \leqslant (\alpha+\beta+\gamma)/C + 12\varepsilon/C$$

Now set $\varepsilon = \rho/C$. We thus obtain

$$2r \leqslant 12\rho/C^2 + C_2$$

where we recall that $C_2 = 3C_1 + 4\varepsilon + 2 = 3\varepsilon(\varepsilon+\rho)/\rho + 4\varepsilon + 2 = \rho(3/C^2 + 7/C) + 2$ with our choice of ε. Since $\rho \geqslant 1$ (unless G is free in which case there is nothing to prove) and necessarily $C \leqslant 1$ we have $7/C \leqslant 7/C^2$ and $2 \leqslant 2\rho/C^2$ and so finally

$$2r \leqslant 12\rho/C^2 + 12\rho/C^2$$

hence the conclusion, recalling that our δ and λ are [Sho91]'s r and ρ respectively. □

2 Growth of random quotients

We now turn to the main point of this paper, namely, evaluation of the growth exponent of a random quotient of a group.

2.1 Framework of the argument

Convention.
Let G_0 be a non-elementary torsion-free hyperbolic group given by the finite presentation $G_0 = \langle\, a_1, \ldots, a_m \mid Q\,\rangle$. Let $g > 0$ be the growth exponent of G_0 with respect to this generating set. Let B_ℓ be the set of elements of norm at most ℓ. Let λ be the maximal length of a relation in Q.

Let also R be a randomly chosen set of $(2m-1)^{d\ell}$ elements of the ball $B_\ell \subset G_0$, in accordance with the model of random quotients we retained (Definition 4). Set $G = G_0/\langle R\rangle$, the random quotient we are interested in. We will call the relators in R "new relators" and those in Q "old relators".

In the sequel, the phrase "with overwhelming probability" will mean "with probability exponentially tending to 1 as $\ell \to \infty$ (depending on everything)".

Fix some $\varepsilon > 0$. We want to show that the growth exponent of G is at least $g(1-\varepsilon)$, with overwhelming probability.

We can suppose that the length ℓ is taken large enough so that, for $L \geqslant \ell$, we have $(2m-1)^{gL} \leqslant |B_L| \leqslant (2m-1)^{g(1+\varepsilon)L}$.

Let $\mathcal{B}_L$ be the ball of radius L in G. We trivially have $|\mathcal{B}_L| \leqslant |B_L|$.

We will prove a lower bound for the cardinality of $\mathcal{B}_L$ for some well chosen L, and then use Proposition 17. In order to apply this proposition, we first need an estimate of the hyperbolicity constant of G.

Proposition 8.
With overwhelming probability, minimal van Kampen diagrams of G satisfy the isoperimetric inequality

$$|\partial D| \geqslant C\,\mathcal{A}(D)$$

where $C > 0$ is a constant depending on G_0 and the density d but not on ℓ. In particular, the hyperbolicity constant δ of G is at most $12\ell/C^2$.

Proof.
This is a rephrasing of Proposition 32 (p. 640) of [Oll04]: With overwhelming probability, minimal van Kampen diagrams D of the random quotient G satisfy the isoperimetric inequality

$$|\partial D| \geqslant \alpha_1 \ell \left|D''\right| + \alpha_2 \left|D'\right|$$

where α_1, α_2 are positive constants depending on G_0 and the density parameter d (but not on ℓ), and $|D''|$, $|D'|$ are respectively the number of faces of D bearing new relators (from R) and old relators (from Q). Since new relators have length at most ℓ and old relators have length at most λ, by definition we have $\mathcal{A}(D) \leqslant \ell\,|D''| + \lambda\,|D'|$ and so setting $C = \min(\alpha_1, \alpha_2/\lambda)$ yields

$$|\partial D| \geqslant C\,\mathcal{A}(D)$$

The estimate of the hyperbolicity constant follows by Proposition 7. □

In particular, in order to apply Proposition 17 it is necessary to control the cardinality of balls of radius roughly $\ell/C^2 + 1/g$. More precisely, let $A \geqslant 500$ be such that $40/A \leqslant \varepsilon/2$. Set $L_0 = 24\ell/C^2 + 4/g$ and $L = AL_0$. We already trivially know that $|\mathcal{B}_{L_0}| \leqslant (2m-1)^{g(1+\varepsilon)L_0}$. We will now show that, with overwhelming probability, we have $|\mathcal{B}_L| \geqslant (2m-1)^{g(1-\varepsilon/2)L}$. Once this is done we can conclude by Proposition 17.

The strategy to evaluate the growth of the quotient G of G_0 will be the following: There are at least $(2m-1)^{gL}$ elements in B_L. Some of these elements are identified in G. Let N be the number of equalities of the form $x = y$, for $x, y \in B_L$, which hold in G but did not hold in G_0. Each such equality decreases the number of elements of $\mathcal{B}_L$ by at most 1. Hence, the number of elements of norm at most L in G is at least $(2m-1)^{gL} - N$. So if we can show for example that $N \leqslant \frac{1}{2}(2m-1)^{gL}$, we will have a lower bound for the size of balls in G.

So we now turn to counting the number of equalities $x = y$ holding in G but not in G_0, with $x, y \in B_L$. Each such equality defines a (minimal) van Kampen diagram with boundary word xy^{-1}, of boundary length at most $2L$. We will need the properties of van Kampen diagrams of G proven in [Oll04].

So, for the ε and A fixed above, let $A' = 2L/\ell$ and let D be a minimal van Kampen diagram of G, of boundary length at most $A'\ell$. By the isoperimetric inequality $|\partial D| \geqslant C\mathcal{A}(D)$, we know that the number $|D''|$ of faces of D bearing a new relator of R is at most A'/C. So for all the sequel set

$$K = A'/C$$

which is the maximal number of new relators in the diagrams we have to consider (which will also have area at most $K\ell$). Most importantly, this K does not depend on ℓ.

2.2 A review of [Oll04]

In this context, it is proven in [Oll04] that the van Kampen diagram D can be seen as a "van Kampen diagram at scale ℓ with respect to the new relators, with equalities modulo G_0". More precisely, this can be stated as follows: (we refer to [Oll04] for the definition of "strongly reduced" diagrams; the only thing to know here is that for any word equal to e in G, there exists a strongly reduced van Kampen diagram with this word as its boundary word).

Proposition 9 ([Oll04], Section 6.6).
Let $G_0 = \langle\, S \mid Q \,\rangle$ be a non-elementary hyperbolic group, let R be a set of words of length ℓ, and consider the group $G = G_0/\langle R\rangle = \langle\, S \mid Q \cup R \,\rangle$.

Let $K \geqslant 1$ be an arbitrarily large integer and let $\varepsilon_1, \varepsilon_2 > 0$ be arbitrarily small numbers. Take ℓ large enough depending on $G_0, K, \varepsilon_1, \varepsilon_2$.

Let D be a van Kampen diagram with respect to the presentation $\langle\, S \mid Q \cup R \,\rangle$, which is strongly reduced, of area at most $K\ell$. Let also D' be the subdiagram of D which is the union of the 1-skeleton of D and of those faces of D bearing relators in Q (so D' is a possibly non-simply connected van Kampen diagram with respect to G_0), and suppose that D' is minimal.

We will call worth-considering *such a van Kampen diagram.*

Let $w_1, \ldots, w_p$ be the boundary (cyclic) words of D', so that each w_i is either the boundary word of D or a relator in R.

Then there exists an integer $k \leqslant 3K/\varepsilon_2$ and words $x_2, \ldots, x_{2k+1}$ such that:

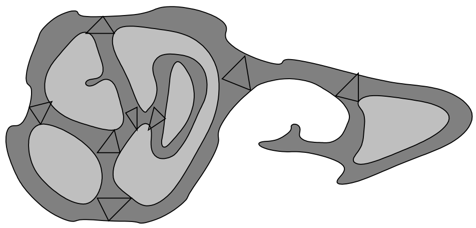
- *Each x_i is a subword of some cyclic word w_j;*
- *As subwords of the w_j's, the x_i's are disjoint and their union exhausts a proportion at least $1 - \varepsilon_1$ of the total length of the w_j's.*
- *For each $i \leqslant k$, there exists words δ_1, δ_2 of length at most $\varepsilon_2(|x_{2i}| + |x_{2i+1}|)$ such that $x_{2i}\delta_1 x_{2i+1}\delta_2 = e$ in G_0.*
- *If two words x_{2i}, x_{2i+1} are subwords of the boundary words of two faces of D bearing the same relator $r^{\pm 1} \in R$, then, as subwords of r, x_{2i} and x_{2i+1} are*

either disjoint or equal with opposite orientations (so that the above equality reads $x\delta_1 x^{-1}\delta_2 = e$).

The pairs (x_{2i}, x_{2i+1}) *are called* translators. *Translators are called* internal, internal-boundary *or* boundary-boundary *according to whether* x_{2i} *and* x_{2i+1} *is a subword of some* w_j *which is a relator in* R *or the boundary word of* D.

(There are slight differences between the presentation here and that in [Oll04]. Therein, boundary-boundary translators did not have to be considered: they were eliminated earlier in the process, before Section 6.6, because they have a positive contribution to boundary length, hence always improve isoperimetry and do not deserve consideration in order to prove hyperbolicity. Moreover, in [Oll04] we further distinguished “commutation translators” for the kind of internal translator with $x_{2i} = x_{2i+1}^{-1}$, which we need not do here.)

Translators appear as dark strips on the following figure:



Remark 10.
Since there are at most $3K/\varepsilon_2$ translators, the total length of the translators (x_{2i}, x_{2i+1}) for which $|x_{2i}| + |x_{2i+1}| \leqslant \varepsilon_3 \ell$ is at most $3K\ell\varepsilon_3/\varepsilon_2$, which makes a proportion at most $3\varepsilon_3/\varepsilon_2$ of the total length. So, setting $\varepsilon_3 = \varepsilon_1\varepsilon_2/3$ and replacing ε_1 with $\varepsilon_1/2$, we can suppose that the union of the translators exhausts a proportion at least $1-\varepsilon_1$ of the total length of the diagram, and that each translator (x_{2i}, x_{2i+1}) satisfies $|x_{2i}| + |x_{2i+1}| \geqslant \varepsilon_1\varepsilon_2\ell/6$.

Remark 11.
The number of ways to partition the words w_i into translators is at most $(2K\ell)^{12K/\varepsilon_2}$, because each w_i can be determined by its starting- and endpoint, which can be given as numbers between 1 and $2K\ell$ which is an upper bound for the cumulated length of the w_i's (since the area of D is at most $K\ell$). For fixed K and ε_2 this grows subexponentially in ℓ.

Remark 12.
Knowing the words x_i, the number of possibilities for the boundary word of the diagram is at most $(6K/\varepsilon_2)!$ (choose which subwords x_i make the boundary word of the diagram, in which order), which does not depend on ℓ for fixed K and ε_2.

We need another notion from [Oll04], namely, that of *apparent length* of an element in G_0. Apparent length is defined in [Oll04] in a more general setting, with respect to a family of measures on the group depending on the precise model of random quotient at play. Here these are simply the uniform measures on the balls B_ℓ. So we only give here what the definition amounts to in our context. In fact we will not use here the full strength of this notion, but we still need to define it in order to state results from [Oll04].

Recall that in the geodesic model of random quotients, the axioms of [Oll04] are satisfied with $\beta = g/2$ and $\kappa_2 = 1$, by Proposition 20 of [Oll04].

Definition 13 ([Oll04], p. 652).
Let $x \in G_0$. Let $\varepsilon_2 > 0$. Let L be an integer. Let $p_L(xuyv = e)$ be the probability that, for a random element $y \in B_L$, there exist elements $u, v \in G_0$ of norm at most $\varepsilon_2(\|x\| + L)$ such that $xuyv = e$ in G_0.

The apparent length *of x at test-length L is*

$$\mathbb{L}_L(x) = -\frac{2}{g} \log_{2m-1} p_L(xuyv = e) - L$$

The apparent length *of x is*

$$\mathbb{L}(x) = \min \left(\|x\| \ , \ \min_{0 \leqslant L \leqslant K\ell} \mathbb{L}_L(x) \right)$$

where we recall ℓ is the length of the relators in a random presentation.

We further need the notion of a *decorated abstract van Kampen diagram* (which was implicitly present in the free case when we mentioned the probability that some diagram "is fulfilled by random relators"), which is inspired by Proposition 9: it carries the combinatorial information about how the relators and boundary word of a diagram were cut into subwords in order to make the translators.

Definition 14 (Decorated abstract van Kampen diagram).
Let $K \geqslant 1$ be an arbitrarily large integer and let $\varepsilon_1, \varepsilon_2 > 0$ be arbitrarily small numbers. Let I_ℓ be the cyclically ordered set of ℓ elements.

A decorated abstract van Kampen diagram *$\mathcal{D}$ is the following data:*

- *An integer $|\mathcal{D}| \leqslant K$ called its* number of faces.
- *An integer $|\partial\mathcal{D}| \leqslant K\ell$ called its* boundary length.
- *An integer $n \leqslant |\mathcal{D}|$ called its* number of distinct relators.
- *An application $r^{\mathcal{D}}$ from $\{1, \ldots, |\mathcal{D}|\}$ to $\{1, \ldots, n\}$; if $r^{\mathcal{D}}(i) = r^{\mathcal{D}}(j)$ we will say that* faces i and j bear the same relator.
- *An integer $k \leqslant 3K/\varepsilon_2$ called the* number of translators *of $\mathcal{D}$.*
- *For each integer $2 \leqslant i \leqslant 2k+1$, a set of the form $\{j_i\} \times I'_i$ where either j_i is an integer between 1 and $|\mathcal{D}|$ and I'_i is an oriented cyclic subinterval of I_ℓ, or $j_i = |\mathcal{D}| + 1$ and I'_i is a subinterval of $I_{|\partial\mathcal{D}|}$; this is called an* (internal) subword of the j_i-th face *in the first case, or a* boundary subword *in the second case.*

- *For each integer* $1 \leqslant i \leqslant k$ *such that* $j_{2i} \leqslant |\mathcal{D}|$*, an integer between* 0 *and* 4ℓ *called the* apparent length of the $2i$-th subword.

such that

- *The sets* $\{j_i\} \times I'_i$ *are all disjoint and the cardinal of their union is at least* $(1 - \varepsilon_1)\left(|\mathcal{D}|\,\ell + |\partial\mathcal{D}|\right)$.
- *For all* $1 \leqslant i \leqslant k$ *we have* $j_{2i} \leqslant j_{2i+1}$ *(this can be ensured by maybe swapping them).*
- *If two faces* j_{2i} *and* j_{2i+1} *bear the same relator, then either* I'_{2i} *and* I'_{2i+1} *are disjoint or are equal with opposite orientations.*

This way, Proposition 9 ensures that any worth-considering van Kampen diagram D with respect to $G_0/\langle R\rangle$ defines a decorated abstract van Kampen diagram $\mathcal{D}$ in the way suggested by terminology (up to rounding the apparent lengths to the nearest integer; we neglect this problem). We will say that $\mathcal{D}$ is *associated to* D. Remark 11 tells that the number of decorated abstract van Kampen diagrams grows subexponentially with ℓ (for fixed K).

Given a decorated abstract van Kampen diagram $\mathcal{D}$ and n given relators $r_1, \ldots, r_n$, we say that these relators *fulfill* $\mathcal{D}$ if there exists a worth-considering van Kampen diagram D with respect to $G_0/\langle r_1, \ldots, r_n\rangle$, such that the associated decorated abstract van Kampen diagram is $\mathcal{D}$. Intuitively speaking, the relators $r_1, \ldots, r_n$ can be "glued modulo G_0 in the way described by $\mathcal{D}$".

So we want to study which diagrams can probably be fulfilled by random relators in R. The main conclusion from [Oll04] is that these are those with large boundary length, hence hyperbolicity of the quotient $G_0/\langle R\rangle$. Here for growth we are rather interested in the number of different elements of G_0 that can appear as boundary words of fulfillable a abstract diagrams with given boundary length (recall that our goal is to evaluate the number of equalities $x = y$ holding in G but not in G_0, with x and y elements of norm at most L).

2.3 Evaluation of growth

We now turn back to random quotients: R is a set of $(2m-1)^{d\ell}$ randomly chosen elements of B_ℓ. Recall we set $L = A'\ell/2$ for some value of A' ensuring that if we know that $|\mathcal{B}_L| \geqslant (2m-1)^{g(1-\varepsilon/2)L}$ then we know that the growth exponent of $G = G_0/\langle R\rangle$ is at least $g(1-\varepsilon)$.

We want to get an upper bound for the number N of pairs $x, y \in B_L$ such that $x = y$ in G but $x \neq y$ in G_0. For any such pair there is a worth-considering van Kampen diagram D with boundary word xy^{-1}, of boundary length at most $A'\ell$, with at most $K = A'/C$ new relators, and at least one new relator (otherwise the equality $x = y$ would already occur in G_0). Let $\mathcal{D}$ be the decorated abstract van Kampen diagram associated to D. Note that we have to count the number of different pairs $x, y \in B_L$ and *not* the number of different boundary words of van Kampen diagrams:

since each x and y may have numerous different representations as a word, the latter is higher than the former.

We will show that, with overwhelming probability, we have $N \leqslant \frac{1}{2}(2m-1)^{gL}$.

The up to now free parameters ε_1 and ε_2 (in the definitions of decorated abstract van Kampen diagrams and of apparent length) will be fixed in the course of the proof, depending on G_0, g and d but not on ℓ. The length ℓ upon which our argument works will be set depending on everything including ε_1 and ε_2.

Further notations. Let n be the number of distinct relators in $\mathcal{D}$. We only have to consider van Kampen diagrams in G which were not already van Kampen diagrams in G_0, so that there is at least one new relator i.e. $n \geqslant 1$. For $1 \leqslant a \leqslant n$, let m_a be the number of times the a-th relator appears in $\mathcal{D}$. Up to reordering, we can suppose that the m_a 's are non-increasing. Also to avoid trivialities take n minimal so that $m_n \geqslant 1$.

Let also P_a be the probability that, if a words $r_1, \ldots, r_a$ of length ℓ are picked at random, there exist $n-a$ words $r_{a+1}, \ldots, r_n$ of lengt ℓ such that the relators $r_1, \ldots, r_n$ fulfill $\mathcal{D}$. The P_a 's are of course a non-increasing sequence of probabilities. In particular, P_n is the probability that a random n-tuple of relators fulfills $\mathcal{D}$.

Back to our set R of $(2m-1)^{d\ell}$ randomly chosen relators. Let P^a be the probability that there exist a relators $r_1, \ldots, r_a$ in R, such that there exist words $r_{a+1}, \ldots, r_n$ of length ℓ such that the relators $r_1, \ldots, r_n$ fulfill $\mathcal{D}$. Again the P^a 's are a non-increasing sequence of probabilities and of course we have

$$P^a \leqslant (2m-1)^{ad\ell} P_a$$

since the $(2m-1)^{ad\ell}$ factor accounts for the choice of the a-tuple of relators in R.

The probability that there exists a van Kampen diagram D with respect to the random presentation R, such that $\mathcal{D}$ is associated to D, is by definition less than P^a for any a. In particular, if for some $\mathcal{D}$ we have $P^a \leqslant (2m-1)^{-\varepsilon'\ell}$, then with overwhelming probability, $\mathcal{D}$ is not associated to any van Kampen diagram of the random presentation. Since, by Remark 11, the number of possibilities for $\mathcal{D}$ grows subexponentially with ℓ, we can sum this over $\mathcal{D}$ and conclude that for any $\varepsilon' > 0$, with overwhelming probability (depending on ε'), all decorated abstract van Kampen diagrams $\mathcal{D}$ associated to some van Kampen diagram of the random presentation satisfy $P^a \geqslant (2m-1)^{-\varepsilon'\ell}$ and in particular

$$P_a \geqslant (2m-1)^{-ad\ell-\varepsilon'\ell}$$

which we assume from now on.

We need to define one further quantity. Keep the notations of Definition 14. Let $1 \leqslant a \leqslant n$ and let $1 \leqslant i \leqslant k$ where k is the number of translators of $\mathcal{D}$. Say that the i-th translator is half finished at time a if $r^{\mathcal{D}}(j_{2i}) \leqslant a$ and $r^{\mathcal{D}}(j_{2i+1}) > a$, that is, if one side of the translator is a subword of a relator $r_{a'}$ with $a' \leqslant a$ and the other of $r_{a''}$ with $a'' > a$. Now let A_a be the sum of the apparent lengths of all translators which are half finished at time a. In particular, A_n is the sum of the apparent lengths of all subwords $2i$ such that $2i$ is an internal subword and $2i+1$ is a boundary subword of $\mathcal{D}$.

Proof of Theorem 2. We first give some intermediate results.

Proposition 15.
With overwhelming probability, we can suppose that any decorated abstract van Kampen diagram $\mathcal{D}$ satisfies

$$A_n(\mathcal{D}) \geqslant \ell\alpha/g + \frac{2}{g}\left(d'_n(\mathcal{D}) + nd\ell\right)$$

where $\alpha = g/2 - d > 0$ and $d'_a(\mathcal{D}) = \log_{2m-1} P_a(\mathcal{D})$.

Proof.
In our context, equation ($\star$) (p. 659) of [Oll04] reads

$$A_a - A_{a-1} \geqslant m_a\left(\ell(1-\varepsilon'') + \frac{\log_{2m-1} P_a - \log_{2m-1} P_{a-1}}{\beta}\right)$$

where ε'' tends to 0 when our free parameters $\varepsilon_1, \varepsilon_2$ tend to 0 (and ε'' also absorbs the $o(\ell)$ term in [Oll04]). Also recall that in the model of random quotient by random elements of balls we have

$$\beta = g/2$$

by Proposition 20 (p. 628) of [Oll04].

Summing over a we get, using $\sum m_a = |\mathcal{D}|$, that

$$\begin{aligned} A_n &\geqslant \left(\sum m_a\right)\ell\left(1-\varepsilon''\right) + \frac{2}{g}\sum m_a(d'_a - d'_{a-1}) \\ &= |\mathcal{D}|\,\ell(1-\varepsilon'') + \frac{2}{g}\sum d'_a(m_a - m_{a+1}) \end{aligned}$$

Now recall we saw above that for any $\varepsilon' > 0$, taking ℓ large enough we can suppose that $P_a \geqslant (2m-1)^{-ad\ell-\varepsilon'\ell}$, that is, $d'_a + ad\ell + \varepsilon'\ell \geqslant 0$. Hence

$$\begin{aligned} A_n &\geqslant |\mathcal{D}|\,\ell(1-\varepsilon'') + \frac{2}{g}\sum(d'_a + ad\ell + \varepsilon'\ell)(m_a - m_{a+1}) \\ &\quad -\frac{2}{g}\sum(ad\ell + \varepsilon'\ell)(m_a - m_{a+1}) \\ &= |\mathcal{D}|\,\ell(1-\varepsilon'') + \frac{2}{g}\sum(d'_a + ad\ell + \varepsilon'\ell)(m_a - m_{a+1}) - \frac{d\ell}{g/2}\sum m_a - \frac{\varepsilon'\ell}{g/2}m_1 \\ &\geqslant |\mathcal{D}|\,\ell(1-\varepsilon'') + \frac{d'_n + nd\ell + \varepsilon'\ell}{g/2}m_n - \frac{d\ell + \varepsilon'\ell}{g/2}\sum m_a \end{aligned}$$

where the last inequality follows from the fact that we chose the order of the relators so that $m_a - m_{a+1} \geqslant 0$.

So using $m_n \geqslant 1$ we finally get

$$A_n \geqslant |\mathcal{D}|\,\ell\left(1-\varepsilon''-\frac{d+\varepsilon'}{g/2}\right)+\frac{d'_n+nd\ell}{g/2}$$

Set $\alpha = g/2 - d > 0$ so that this rewrites

$$A_n \geqslant \frac{2}{g}\left(|\mathcal{D}|\,\ell\left(\alpha-\varepsilon'-\varepsilon'' g/2\right)+d'_n+nd\ell\right)$$

Suppose the free parameters ε_1, ε_2 and ε' are chosen small enough so that $\varepsilon' + \varepsilon'' g/2 \leqslant \alpha/2$ (recall that ε'' is a function of $\varepsilon_1, \varepsilon_2$ and K, tending to 0 when ε_1 and ε_2 tend to 0). Since $|\mathcal{D}| \geqslant 1$ (because we are counting diagrams expressing equalities not holding in G_0) we get $A_n \geqslant \ell\alpha/g + \frac{2}{g}\left(d'_n + nd\ell\right)$. □

Let us translate back this inequality into a control on the numbers of n-tuples of relators fulfilling $\mathcal{D}$.

Proposition 16.
With overwhelming probability, we can suppose that for any decorated abstract van Kampen diagram $\mathcal{D}$, the number of n-tuples of relators in R fulfilling $\mathcal{D}$ is at most

$$(2m-1)^{-\alpha\ell/2+gA_n(\mathcal{D})/2+\varepsilon'\ell}$$

Proof.
Recall that, by definition, d'_n is the log-probability that n random relators $r_1, \ldots, r_n$ fulfill $\mathcal{D}$. As there are $(2m-1)^{nd\ell}$ n-tuples of random relators in R (by definition of the density model), by linearity of expectation the expected number of n-tuples of relators in R fulfilling $\mathcal{D}$ is $(2m-1)^{nd\ell+d'_n}$.

By the Markov inequality, for given $\mathcal{D}$ the probability to pick a random set R such that the number of n-tuples of relators of R fulfilling $\mathcal{D}$ is greater than $(2m-1)^{nd\ell+d'_n+\varepsilon'\ell}$, is less than $(2m-1)^{-\varepsilon'\ell}$. Using Proposition 15, the result then follows for fixed $\mathcal{D}$. But by Remark 11 the number of possibilities for $\mathcal{D}$ is subexponential in ℓ, hence the conclusion. □

Let us now turn back to the evaluation of the number of elements x, y in $B_L \subset G_0$ forming a van Kampen diagram D with boundary word xy^{-1}. For each such pair x, y fix some geodesic writing of x and y as words. We will first suppose that the abstract diagram $\mathcal{D}$ associated to D is fixed and evaluate the number of possible pairs x, y in function of $\mathcal{D}$, and then, sum over the possible abstract diagrams $\mathcal{D}$.

So suppose $\mathcal{D}$ is fixed. Recall Proposition 9: the boundary word of D is determined by giving two words for each boundary-boundary translator, and one word for each internal-boundary translator, this last one being subject to the apparent length condition imposed in the definition of $\mathcal{D}$. By Remark 12, the number of ways to combine these subwords into a boundary word for D is controlled by K and ε_2 (independently of ℓ).

In all the sequel, in order to avoid heavy notations, the notation $\varepsilon^\star$ will denote some function of ε', ε_1 and ε_2, varying from time to time, and increasing when needed. The important point is that $\varepsilon^\star$ tends to 0 when ε', ε_1, ε_2 do.

Let (x_{2i}, x_{2i+1}) be a translator in D. The definition of translators implies that there exist short words δ_1, δ_2, of length at most $\varepsilon_2(|x_{2i}|+|x_{2i+1}|)$, such that $x_{2i}\delta_1 x_{2i+1}\delta_2 = e$ in G_0. The words x_{2i} and x_{2i+1} are either subwords of the geodesic words x and y making the boundary of D, or subwords of relators in R; by definition of the geodesic model of random quotients, the relators are geodesic as well. So in either case x_{2i} and x_{2i+1} are geodesic[1]. Thus, the equality $x_{2i}\delta_1 x_{2i+1}\delta_2 = e$ implies that $\|x_{2i+1}\| \leqslant \|x_{2i}\|(1+\varepsilon^\star)$ and conversely. Also, by Remark 10, we can suppose that $\|x_{2i}\| + \|x_{2i+1}\| \geqslant \ell\varepsilon_1\varepsilon_2/6$, hence $\|x_{2i}\| \geqslant \ell\varepsilon_1\varepsilon_2(1-\varepsilon^\star)/12$.

By definition of the growth exponent, there is some length ℓ_0 depending only on G_0 such that if $\ell'_0 \geqslant \ell_0$, then the cardinal of $B_{\ell'_0}$ is at most $(2m-1)^{g(1+\varepsilon')\ell'_0}$. So, if ℓ is large enough (depending on G_0, ε_1, ε_2 and ε') to ensure that $\ell\varepsilon_1\varepsilon_2(1-\varepsilon^\star)/12 \geqslant \ell_0$, we can apply such an estimate to any x_{2i}.

To determine the number of possible pairs x, y, we have to determine the number of possibilites for each boundary-boundary or internal-boundary translator (x_{2i}, x_{2i+1}) (since by definition internal translators do not contribute to the boundary).

First suppose that (x_{2i}, x_{2i+1}) is a boundary-boundary translator. Knowing the constraint $x_{2i}\delta_1 x_{2i+1}\delta_2 = e$, if x_{2i} and $\delta_{1,2}$ are given then x_{2i+1} is determined (as an element of G_0). The number of possibilities for δ_1 and δ_2 is at most $(2m-1)^{2\varepsilon_2(\|x_{2i}\|+\|x_{2i+1}\|)}$. The number of possibilities for x_{2i} is at most $(2m-1)^{g(1+\varepsilon')\|x_{2i}\|}$ which, since $\|x_{2i}\| \leqslant \frac{1}{2}(\|x_{2i}\|+\|x_{2i+1}\|)(1+\varepsilon^\star)$, is at most $(2m-1)^{\frac{g}{2}(\|x_{2i}\|+\|x_{2i+1}\|)(1+\varepsilon^\star)}$. So the total number of possibilities for a boundary-boundary translator (x_{2i}, x_{2i+1}) is at most

$$(2m-1)^{\frac{g}{2}(\|x_{2i}\|+\|x_{2i+1}\|)(1+\varepsilon^\star)}$$

where of course the feature to remember is that the exponent is basically $g/2$ times the total length $\|x_{2i}\|+\|x_{2i+1}\|$ of the translator.

Now suppose that (x_{2i}, x_{2i+1}) is an internal-boundary translator. The word x_{2i} is by definition a subword of some relator $r_i \in R$. So if a set of relators fulfilling $\mathcal{D}$ is fixed then x_{2i} is determined (we will multiply later by the number of possibilities for the relators, using Proposition 16). As above, the number of possibilities for δ_1 and δ_2 is at most $(2m-1)^{\varepsilon^\star\|x_{2i}\|}$. Once x_{2i}, δ_1 and δ_2 are given, then x_{2i+1} is determined (as an element of G_0). So, if a set of relators fulfilling $\mathcal{D}$ is fixed, then the number of possibilities for x_{2i+1} is at most $(2m-1)^{\varepsilon^\star\|x_{2i}\|}$, which reflects the fact that the set of relators essentially determines the internal-boundary translators.

Let A'_n be the sum of $\|x_{2i+1}\|$ for all internal-boundary translators (x_{2i}, x_{2i+1}). Let B be the sum of $\|x_{2i}\|+\|x_{2i+1}\|$ for all boundary-boundary translators. By definition we have $|\partial\mathcal{D}| = A'_n + B$ maybe up to $\varepsilon_1 K\ell$.

[1]Except maybe in the case when the translator straddles the end of x and the beginning of y or conversely, or when it straddles the beginning and end of a relator; these cases can be treated immediately by further subdividing the translator, so we ignore this problem.

So if a set of relators fulfilling $\mathcal{D}$ is fixed, then the total number of possibilities for the boundary of D is at most

$$(2m-1)^{\frac{g}{2}B(1+\varepsilon^\star)+\varepsilon^\star A'_n}$$

which, since both B and A'_n are at most $K\ell$, is at most

$$(2m-1)^{gB/2+K\ell\varepsilon^\star}$$

(note that A'_n does not come into play, since once the relators fulfilling $\mathcal{D}$ are given, the internal-boundary translators are essentially determined).

The number of possibilities for an n-tuple of relators fulfilling $\mathcal{D}$ is given by Proposition 16: it is at most $(2m-1)^{-\alpha\ell/2+gA_n/2+\varepsilon^\star\ell}$ (recall $\alpha = g/2-d$), so that the total number of possibilities for the boundary of D is at most

$$(2m-1)^{-\alpha\ell/2+(B+A_n)g/2+K\ell\varepsilon^\star}$$

Recall that A_n is the sum of $\mathbb{L}(x_{2i})$ for all internal-boundary translators (x_{2i}, x_{2i+1}). By definition of apparent length we have $\mathbb{L}(x_{2i}) \leqslant \|x_{2i}\|$. Since in an internal-boundary translator (x_{2i}, x_{2i+1}) we have $\|x_{2i}\| \leqslant \|x_{2i+1}\|(1+\varepsilon^\star)$, we get, after summing on all internal-boundary translators, that $A_n \leqslant A'_n + K\ell\varepsilon^\star$. In particular, the above is at most

$$(2m-1)^{-\alpha\ell/2+(B+A'_n)g/2+K\ell\varepsilon^\star}$$

Now recall that by definition we have $|\partial\mathcal{D}| = B + A'_n$ maybe up to $\varepsilon_1 K\ell$ so that the above is in turn at most

$$(2m-1)^{-\alpha\ell/2+|\partial\mathcal{D}|g/2+K\varepsilon^\star\ell}$$

This was for one decorated abstract van Kampen diagram $\mathcal{D}$. But by Remark 11, the number of such diagrams is subexponential in ℓ (for fixed K and ε_2), and so, up to increasing $\varepsilon^\star$, this estimate holds for all diagrams simultaneously.

2.4 Conclusion

Remember the discussion in the beginning of Section 2. We wanted to show that the cardinal $|\mathcal{B}_L|$ of the ball of radius L in G was at least $(2m-1)^{gL(1-\varepsilon/2)}$ for some ε chosen at the beginning of our work.

We just proved that the number N of pairs of elements x, y in B_L such that there exists a van Kampen diagram expressing the equality $x = y$ in G, but such that $x \neq y$ in G_0 (which was expressed in the above argument by using that D had at least one new relator) is at most

$$(2m-1)^{-\alpha\ell/2+(\|x\|+\|y\|)g/2+K\varepsilon^\star\ell}$$

where $\alpha = g/2 - d > 0$.

Now fix the free parameters ε', ε_1, ε_2 so that $K\varepsilon^\star \leqslant \alpha/4$ (this depends on K and G_0 but not on ℓ; K itself depends only on G_0). Choose ℓ large enough so that all the estimates used above (implying every other variable) hold. Also choose ℓ large enough (depending on d) so that $(2m-1)^{-\alpha\ell/4} \leqslant 1/2$. We get

$$N \leqslant \frac{1}{2}(2m-1)^{(\|x\|+\|y\|)g/2} \leqslant \frac{1}{2}(2m-1)^{gL}$$

since by assumption $\|x\|$ and $\|y\|$ are at most L. But on the other hand we have $|B_L| \geqslant (2m-1)^{gL}$ and so

$$|\mathcal{B}_L| \geqslant |B_L| - N \geqslant \frac{1}{2}(2m-1)^{gL} \geqslant (2m-1)^{gL(1-\varepsilon/2)}$$

as soon as ℓ is large enough (since L grows like ℓ), which ends the proof.

Appendix: Locality of growth in hyperbolic groups

The goal of this section is to show that, in a hyperbolic group, if we know an estimate of the growth exponent in some finite ball of the group, then this provides an estimate of the growth exponent of the group (whose quality depends on the radius of the given finite ball).

Let $G = \langle\, a_1, \ldots, a_m \mid R\,\rangle$ be a δ-hyperbolic group generated by the elements $a_i^{\pm 1}$, with $m \geqslant 2$. For $x \in G$ let $\|x\|$ be the norm of x with respect to this generating set. Let B_ℓ be the set of elements of norm at most ℓ.

Proposition 17.
Suppose that for some $g > 0$, for some $\ell_0 \geqslant 2\delta + 4/g$ and $\ell_1 \geqslant A\ell_0$, with $A \geqslant 500$, we have

$$|B_{\ell_0}| \leqslant (2m-1)^{1.1g\ell_0}$$

and

$$|B_{\ell_1}| \geqslant (2m-1)^{g\ell_1}$$

Then the growth exponent of G is at least $g(1-40/A)$.

Note that the occurrence of $1/g$ in the scale upon which the proposition is true is natural: indeed, an assumption such as $|B_\ell| \geqslant (2m-1)^{g\ell}$ for $\ell < 1/g$ is not very strong... The growth g can be thought of as the inverse of a length, so this result is homogeneous.

Corollary 18.
The growth exponent of a presentation of a hyperbolic group is computable. That is, there exists an algorithm which, for any input made of a finite presentation of a hyperbolic group and an $\varepsilon > 0$, outputs a number g together with a proof that the growth exponent of the given presentation lies between $g - \varepsilon$ and $g + \varepsilon$.

This corollary was already known: indeed, once δ is known one can compute (see [GhH90]) a finite automaton accepting some normal geodesic form of all elements in the group, and this in turn implies that the growth series is a rational function with explicitly computable coefficients; now the growth exponent is linked to the radius of convergence of this series, which is computable in the case of a rational function. Whereas in this approach, the exact value of the growth exponent is determined very indirectly by the full algebraic structure of some finite ball, our approach directly relates an approximate value of the growth exponent to that observed in this finite ball.

Proof.
Indeed, recall from [Pap96] (after [Gro87]) that the hyperbolicity constant δ of a presentation of a hyperbolic group is computable. Thanks to the isoperimetric inequality, the word problem in a hyperbolic group is solvable, so that for any ℓ an exact computation of the cardinal of B_ℓ is possible. Setting $g_\ell = \frac{1}{\ell}\log_{2m-1}|B_\ell|$, we know that g_ℓ will converge to some (unknown) positive value, so that g_ℓ and $g_{A\ell}$ will become arbitrarily close, and since g_ℓ is bounded from below sooner or later we will have $\ell \geqslant 2\delta + 4/g_{A\ell}$, in which case we can apply the proposition to ℓ and $A\ell$. □

Proof of the proposition.
Let $(,)$ denote the Gromov product in G, with origin at e, that is

$$(x,y) = \frac{1}{2}\left(\|x\| + \|y\| - \|x-y\|\right)$$

for $x, y \in G$, where, following [GhH90], we write $\|x-y\|$ for $\left\|x^{-1}y\right\| = \left\|y^{-1}x\right\|$. Since triangles are δ-thin, we have ([GhH90], Proposition 2.21) for any three points x,y, z in G

$$(x,z) \geqslant \min\left((x,y),(y,z)\right) - 2\delta$$

Let S_ℓ denote the set of elements of norm ℓ in the hyperbolic group G. Consider also, for homogeneity reasons, the annulus $S_{\ell,a} = B_\ell \setminus B_{\ell-a}$.

Proposition 19.
Let $g \in B_\ell$ and let $a \geqslant 0$. The number of elements g' in S_ℓ or B_ℓ such that $(g,g') \geqslant a$ is at most $|B_{\ell-a+2\delta}|$.

Proof.
Suppose that $(g,g') \geqslant a$. Let x be the point at distance a from e on some geodesic joining e to g. By construction we have $(g,x) = a$. But

$$(g',x) \geqslant \min\big((g',g),(g,x)\big) - 2\delta \geqslant a - 2\delta$$

and unwinding the definition of (g',x) yields

$$\left\|g'-x\right\| \leqslant \left\|g'\right\| + \|x\| - 2a + 2\delta \leqslant \ell - a + 2\delta$$

So g' lies at distance at most $\ell - a + 2\delta$ from x, hence the number of possibilities for g' is at most $|B_{\ell-a+2\delta}|$. (This is most clear on a picture.) □

We know show that, if we multiply two elements of the sphere S_ℓ then we often get an element of norm close to 2ℓ.

Corollary 20.
Let $g \in S_{\ell,a}$. The number of elements g' in $S_{\ell,a}$ such that $\|gg'\| \geqslant 2\ell - 4a$ is at least $|S_{\ell,a}| - |B_{\ell-a+2\delta}|$.

Proof.
We have $\|gg'\| = \|g\| + \|g'\| - 2(g^{-1}, g')$. So if $\|g\| \geqslant \ell - a$, $\|g'\| \geqslant \ell - a$ and $(g^{-1}, g') \leqslant a$, then $\|gg'\| \geqslant 2\ell - 4a$.

But by the last proposition, the number of "bad" elements g' such that $(g^{-1}, g') \geqslant a$ is at most $|B_{\ell-a+2\delta}|$. □

So multiplying long elements often gives twice as long elements. We now show that this procedure does not build too often the same new element.

Proposition 21.
Let $x \in S_{2\ell,4a}$. The number of pairs (g, g') in $S_{\ell,a} \times S_{\ell,a}$ such that $x = gg'$ is at most $|B_{6a+2\delta}|$.

Proof.
Choose a geodesic decomposition $x = hh'$ with $\|h\| = \|h'\| = \|x\|/2$. It is easy to see that if $x = gg'$ as above, then g is $6a + 2\delta$-close to h (and then g' is determined). □

Combining the last two results yields the following "almost supermultiplicative" estimate for the cardinals of balls (compare the trivial converse inequality $|B_{2\ell}| \leqslant |B_\ell|^2$).

Corollary 22.

$$|B_{2\ell}| \geqslant \frac{1}{|B_{6a+2\delta}|} \left(|B_\ell| - 2\,|B_{\ell-a+2\delta}|\right)^2$$

Proof.
Indeed, the last two results imply that

$$|S_{2\ell,4a}| \geqslant \frac{1}{|B_{6a+2\delta}|} |S_{\ell,a}| \left(|S_{\ell,a}| - |B_{\ell-a+2\delta}|\right)$$

which implies the above by the trivial estimates $|B_{2\ell}| \geqslant |S_{2\ell,4a}|$ and $|S_{\ell,a}| \geqslant |B_\ell| - |B_{\ell-a+2\delta}|$. □

In order to apply this, we need to know both that $|B_\ell|$ is large and that $|B_{\ell-a}|$ is not too large compared to $|B_\ell|$. Asymptotically one would expect $|B_{\ell-a}| \approx (2m-1)^{-ga} |B_\ell|$. The next lemma states that, under the assumptions of Proposition 17, we can almost realize this, up to changing ℓ by some controlled factor.

Lemma 23.
Suppose that for some g, for some ℓ_0 and $\ell_1 \geqslant 100\ell_0$ we have $|B_{\ell_0}| \leqslant (2m-1)^{1.2g\ell_0}$ and $|B_{\ell_1}| \geqslant (2m-1)^{g\ell_1}$. Let $a \leqslant \ell_0$. There exists $0.65\ell_1 \leqslant \ell \leqslant \ell_1$ such that

$$|B_\ell| \geqslant (2m-1)^{g\ell}$$

and

$$|B_\ell| \geqslant (2m-1)^{ga/2}\,|B_{\ell-a}|$$

Proof of the lemma.
First, note that by subadditivity, the inequality $|B_{\ell_0}| \leqslant (2m-1)^{1.2g\ell_0}$ implies that for any ℓ, writing $\ell = k\ell_0 - r$ ($k \in \mathbb{N}, 0 \leqslant r < \ell_0$) we have $|B_\ell| \leqslant (2m-1)^{1.2kg\ell_0}$. Especially for $\ell \geqslant 50\ell_0$ we have $1 \leqslant k\ell_0/\ell \leqslant 51/50$ and so in particular, if $\ell_1 \geqslant 100\ell_0$ then $|B_{0.65\ell_1}| \leqslant (2m-1)^{0.8g\ell_1}$ (indeed $0.65 \times 1.2 \times 51/50 \leqslant 0.8$).

Suppose that for all $0.65\ell_1 \leqslant \ell \leqslant \ell_1$ with $\ell = \ell_1 - ka$ ($k \in \mathbb{N}$) we have $|B_\ell| < (2m-1)^{ga/2}\,|B_{\ell-a}|$. Write $\ell_1 - 0.65\ell_1 = qa - r$ with $q \in \mathbb{N}$, $0 \leqslant r < a$. Then we get

$$\begin{aligned}
|B_{\ell_1}| &< (2m-1)^{ga/2}\,|B_{\ell_1-a}| < (2m-1)^{ga}\,|B_{\ell_1-2a}| < \cdots \\
&< (2m-1)^{gqa/2}\,|B_{0.65\ell_1-r}| \leqslant (2m-1)^{g(\ell_1-0.65\ell_1)/2+ga/2}\,|B_{0.65\ell_1}| \\
&\leqslant (2m-1)^{g(0.35\ell_1)/2+g\ell_1/200+0.8g\ell_1} < (2m-1)^{0.98g\ell_1}
\end{aligned}$$

contradicting the assumption.

So we can safely take the largest $\ell \leqslant \ell_1$ satisfying $|B_\ell| \geqslant (2m-1)^{ga/2}\,|B_{\ell-a}|$ and such that $\ell_1 - \ell$ is a multiple of a.

Since ℓ is largest, for $\ell \leqslant \ell' \leqslant \ell_1$ we have $|B_{\ell'}| \leqslant (2m-1)^{ga/2}\,|B_{\ell'-a}|$. We get, a-step by a-step, that $|B_{\ell_1}| \leqslant (2m-1)^{g(\ell_1-\ell)/2}\,|B_\ell|$. Using the assumption $|B_{\ell_1}| \geqslant (2m-1)^{g\ell_1}$ we now get $|B_\ell| \geqslant (2m-1)^{g\ell_1 - g(\ell_1-\ell)/2} \geqslant (2m-1)^{g\ell}$ as needed. □

Now equipped with the lemma, we can apply Corollary 22 to show that if we know that B_ℓ is large for some ℓ, then we get a larger ℓ' such that $B_{\ell'}$ is large as well. We will then conclude by induction.

Lemma 24.
Suppose that for some g, for some $\ell_0 \geqslant 2\delta + 4/g$ and $\ell_1 \geqslant A\ell_0$ (with $A \geqslant 100$) we have $|B_{\ell_0}| \leqslant (2m-1)^{1.2g\ell_0}$ and $|B_{\ell_1}| \geqslant (2m-1)^{g\ell_1}$. Then there exists $\ell_2 \geqslant 1.3\ell_1$ such that

$$|B_{\ell_2}| \geqslant (2m-1)^{g\ell_2(1-9/A)}$$

Proof of the lemma.
Consider the ℓ provided by Lemma 23 where we take $a = \ell_0$. This provides an $\ell \geqslant 0.65\ell_1$ such that $|B_\ell| \geqslant (2m-1)^{g\ell}$ and $|B_\ell| \geqslant (2m-1)^{ga/2}\,|B_{\ell-a}|$.

So by Corollary 22 (applied to $2a$ instead of a) we have

$$|B_{2\ell}| \geqslant \frac{1}{|B_{12a+2\delta}|}\,|B_\ell|^2\,(1 - 2\,|B_{\ell-2a+2\delta}|\,/\,|B_\ell|)^2$$

Since $a = \ell_0 \geqslant 2\delta$ we have $\ell - 2a + 2\delta \leqslant \ell - \ell_0$ and so

$$|B_{2\ell}| \geqslant \frac{1}{|B_{12\ell_0+2\delta}|}\,|B_\ell|^2\left(1 - 2(2m-1)^{-g\ell_0/2}\right)^2$$

If $\ell_0 \geqslant 4/g$, since $2m-1 \geqslant 2$ we have $\left(1-2(2m-1)^{-g\ell_0/2}\right)^2 \geqslant 1/4$ and so

$$|B_{2\ell}| \geqslant \frac{1}{4\,|B_{12\ell_0+2\delta}|}\,|B_\ell|^2$$

We have $|B_{12\ell_0+2\delta}| \leqslant |B_{13\ell_0}| \leqslant |B_{\ell_0}|^{13}$ by subadditivity. So by the assumptions

$$|B_{2\ell}| \geqslant \frac{1}{4\,|B_{\ell_0}|^{13}}\,|B_\ell|^2 \geqslant (2m-1)^{2g\ell-16g\ell_0-2} = (2m-1)^{2g\ell(1-8\ell_0/\ell-1/g\ell)}$$

which is at least $(2m-1)^{2g\ell(1-9/A)}$ since $8\ell_0/\ell \leqslant 8/A$ and $1/g\ell \leqslant 1/gA\ell_0 \leqslant 1/A$ since $\ell_0 \geqslant 4/g$.

So we can take $\ell_2 = 2\ell$, which is at least $1.3\ell_1$. □

Now the proposition is clear: start from ℓ_1 and construct by induction a sequence ℓ_i with $\ell_{i+1} \geqslant 1.3\ell_i$ using the lemma applied to ℓ_0 and ℓ_i; thus

$$|B_{\ell_i}| \geqslant (2m-1)^{g\ell_i \prod_{k=0}^{i-2}(1-9/(A\cdot 1.3^k))}$$

and note that the infinite product converges to a value greater than $1-40/A$. The only thing to check is that, in order to be allowed to apply the previous lemma to ℓ_0 and ℓ_i at each step, we must ensure that $1.1/(1-40/A) \leqslant 1.2$, which is guaranteed as soon as $A \geqslant 500$. □

References

[AL02] G.N. Arzhantseva, I.G. Lysenok, *Growth tightness for word hyperbolic groups*, Math. Z. **241** (2002), n° 3, 597–611.

[Ch91] C. Champetier, *Propriétés génériques des groupes de présentation finie*, Ph.D. Thesis, Université Lyon 1 (1991).

[Ch93] C. Champetier, *Cocroissance des groupes à petite simplification*, Bull. London Math. Soc. **25** (1993), n° 5, 438–444.

[Ch95] C. Champetier, *Propriétés statistiques des groupes de présentation finie*, J. Adv. Math. **116** (1995), n° 2, 197–262.

[Gh03] É. Ghys, *Groupes aléatoires*, séminaire Bourbaki **916** (2003).

[GhH90] É. Ghys, P. de la Harpe, *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math. **83**, Birkhäuser (1990).

[GrH97] R.I. Grigorchuk, P. de la Harpe, *On problems related to growth, entropy, and spectrum in group theory*, Dynam. Control Systems **3** (1997), n° 1, 51–89.

[Gro87] M. Gromov, *Hyperbolic Groups*, in *Essays in group theory*, ed. S.M. Gersten, Springer (1987), 75–265.

[Gro93] M. Gromov, *Asymptotic Invariants of Infinite Groups*, in *Geometric group theory*, ed. G. Niblo, M. Roller, Cambridge University Press, Cambridge (1993).

[Gro03] M. Gromov, *Random Walk in Random Groups*, Geom. Funct. Anal. **13** (2003), n° 1, 73–146.

[Har00] P. de la Harpe, *Topics in geometric group theory*, Chicago University Press (2000).

[HLS02] N. Higson, V. Lafforgue, G. Skandalis, *Counterexamples to the Baum-Connes conjecture*, Geom. Funct. Anal. **12** (2002), n° 2, 330–354.

[LS77] R.C. Lyndon, P.E. Schupp, *Combinatorial Group Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete **89**, Springer (1977).

[Oll04] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, GAFA, Geom. Funct. Anal. **14** (2004), n° 3, 595–679.

[Oll05] Y. Ollivier, *Cogrowth and spectral gap of generic groups*, Ann. Inst. Fourier (Grenoble) **55** (2005), n° 1, 289–317.

[Oll-a] Y. Ollivier, *Some small cancellation properties of random groups*, to appear in Internat. J. Algebra Comput., ArXiV `math.GR/0409226`

[Oll-b] Y. Ollivier, *A January 2005 Invitation to Random Groups*, expository manuscript.

[Ols92] A.Yu. Ol'shanskiĭ, *Almost Every Group is Hyperbolic*, Internat. J. Algebra Comput. **2** (1992), n° 1, 1–17.

[Pap96] 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.

[Sho91] H. Short et al., in *Group Theory from a Geometrical Viewpoint*, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991).

[Shu99] A.G. Shukhov, *On the dependence of the growth exponent on the length of the defining relation*, Math. Notes **65** (1999), n° 3–4, 510–515.

[Ver00] A.M. Vershik, *Dynamic theory of growth in groups: entropy, boundaries, examples*, Russian Math. Surveys **55** (2000), n° 4, 667–733.

Habilitation à diriger des recherches

Some small cancellation properties of random groups

*Les idées contenues dans ce texte datent de mon séjour chez Dani Wise à Montréal en août 2004, et sont apparues comme des outils nécessaires, mais d'intérêt indépendant, pour notre travail. Ce texte a été publié dans le volume 17, n°1 (2007) de l'*International Journal of Algebra and Computation*, pp. 37–51.*

Yann Ollivier, École normale supérieure de Lyon, 2009

Some small cancellation properties of random groups

Yann Ollivier

Abstract

We work in the density model of random groups. We prove that they satisfy an isoperimetric inequality with sharp constant $1-2d$ depending upon the density parameter d. This implies in particular a property generalizing the ordinary C' small cancellation condition, which could be termed "macroscopic cancellation". This also sharpens the evaluation of the hyperbolicity constant δ.

As a consequence we get that the standard presentation of a random group at density $d < 1/5$ satisfies the Dehn algorithm and Greendlinger's Lemma, and that it does not for $d > 1/5$.

For this we establish a version of the local-global principle for hyperbolic spaces (Cartan-Hadamard-Gromov theorem) involving arbitrarily small loss in the isoperimetric constant.

Statements

Gromov introduced in [Gro93] the so-called *density model* of random groups, which allows the study of generic groups with a very precise control on the number of relators put in the group, depending on a density parameter d.

A set of m generators $a_1, \ldots, a_m$ being fixed, this model consists in choosing a (large) length ℓ and a density parameter $0 \leqslant d \leqslant 1$, and choosing at random a set R of $(2m-1)^{d\ell}$ reduced words of length ℓ. The random group is then the group given by the presentation $\langle\, a_1, \ldots, a_m \mid R \,\rangle$. (Recall a word is *reduced* if it does not contain a generator immediately followed by its inverse).

In this model, we say that a property occurs *with overwhelming probability* if its probability of occurrence tends to 1 as $\ell \to \infty$ (everything else being fixed).

The basic intuition behind the model is that at density d, subwords of length $(d-\varepsilon)\ell$ of the relators will exhaust all possible reduced words of this length. Also, at density d, with overwhelming probability there are two relators sharing a subword of length $(2d-\varepsilon)\ell$. We refer to [Gro93], [Ghy03] or [Oll-b] for a general discussion on random groups and the density model.

The interest of this way to measure the number of relators in a presentation is largely established by the following foundational theorem, due to Gromov ([Gro93], see also [Oll04]).

Theorem 1 (M. Gromov).
If $d < 1/2$, with overwhelming probability a random group at density d is infinite and hyperbolic.

If $d > 1/2$, with overwhelming probability a random group at density d is either $\{e\}$ or $\mathbb{Z}/2\mathbb{Z}$.

(Occurrence of $\mathbb{Z}/2\mathbb{Z}$ of course corresponds to even ℓ.)

Other properties of random groups are known, some of which depending on the density parameter (works of Arzhantseva, Champetier, Gromov, Ollivier, Ol'shanksiĭ, Żuk; see references in [Ghy03, Oll04, Oll-b]). The construction can be modified and iterated in various ways to achieve specific goals [Gro03].

Hyperbolicity for $d < 1/2$ is achieved by proving that van Kampen diagrams satisfy some isoperimetric inequality (we refer to [LS77] for definitions about van Kampen diagrams and to [Sho91] for the equivalence between hyperbolicity and isoperimetry of van Kampen diagrams). The main result of this paper is a sharp version of this isoperimetric inequality.

Theorem 2.
For every $\varepsilon > 0$, with overwhelming probability, every reduced van Kampen diagram D in a random group at density d satisfies

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

This was already known to hold for diagrams of size bounded a priori (see Theorem 14), but the passage to all diagrams involves the local-global hyperbolic principle of Gromov (see e.g. [Pap96] for a constructive statement), which implies a substantial loss in the constants. After using this, the only constant available for all diagrams was something like $(1 - 2d)/10^{20}$. We solve the problem by giving a variant of the principle best suited to our needs (Theorem 8), which may have independent interest.

This inequality is sharp: indeed, at density d there are very probably two relators sharing a subword of length $(2d - \varepsilon)\ell$, so that they can be arranged to form a 2-face van Kampen diagram of boundary length $2(1 - 2d + \varepsilon)\ell$. At density d one can always glue some new relator to any diagram along a path of length $(d - \varepsilon)\ell$, so that adding relators to this example provides an arbitrarily large diagram with the same isoperimetric constant.

Corollary 3.
At density d, with overwhelming probability the hyperbolicity constant of a random group satisfies $\delta \leqslant 4\ell/(1 - 2d)$.

Corollary 4.
For every $\varepsilon > 0$, with overwhelming probability, random groups at density d satisfy the following: Let D_1 and D_2 be two reduced van Kampen diagrams, both of them

homeomorphic to a disk. Suppose that their boundaries share a common reduced subword w. Suppose moreover that the diagram $D = D_1 \cup_w D_2$ obtained by gluing D_1 and D_2 along w is still reduced. Then we have

$$|w| \leqslant d\,(|\partial D_1| + |\partial D_2|)\,(1+\varepsilon)$$

When D_1 and D_2 each consist of only one face, this exactly states that random groups satisfy the $C'(2d)$ small cancellation property (which implies hyperbolicity only when $d < 1/12$). So this property is a kind of "macroscopic cancellation" (though not "small" cancellation when d is close to $1/2$).

Our last application of Theorem 2 has to do with the Dehn algorithm and Greendlinger's Lemma, which are classical properties considered in combinatorial group theory (see [LS77], [Gre60]).

There are several versions of Greendlinger's Lemma. We will not use the strongest version which holds for $C'(1/6)$ presentations ([LS77], Theorem V.4.5). The exact property we will use is the following.

Given a face f of a van Kampen diagram D, a *contour segment* of f in D is a subset of edges of $\partial f \cap \partial D$ which are consecutive in the boundary path of D.

Definition 5 (Greendlinger's property).
We say that a group presentation satisfies the Greendlinger property *if the following holds: For any reduced van Kampen diagram D w.r.t. the presentation, with reduced boundary word, either D has only one face or there exist at least two faces of D having contour segments of lengths more than half their respective lengths.*

Of course this implies that Dehn's algorithm works.

One might expect from Theorem 2 that the Dehn algorithm holds as soon as $d < 1/4$. Indeed, $d < 1/4$ implies that some face of any reduced diagram has at least $\ell/2$ boundary edges; but these might not be consecutive. Actually the critical density is $1/5$.

Theorem 6.
If $d < 1/5$, with overwhelming probability, the standard presentation of a random group satisfies the Dehn algorithm and the Greendlinger property.

More precisely, for any $\varepsilon > 0$, with overwhelming probability, in every reduced van Kampen diagram with reduced boundary word, with at least two faces, there are at least two faces having a contour segment of length more than $\frac{\ell}{2} + \frac{\ell}{2}(1 - 5d - \varepsilon)$.

If $d > 1/5$, with overwhelming probability, the standard presentation of a random group does not satisfy the Dehn algorithm nor the Greendlinger property.

See p. 281 for a simple example of a van Kampen diagram violating the Greendlinger property when $d > 1/5$.

Discussion of the results. The interest of the sharp constant depending on density in Theorem 2, compared to the 10^{20} times smaller previous estimate, is not only aesthetic. Let us stress that the Dehn algorithm could not be obtained with the previous constant, if only for the reason that $(1-2d)/10^{20}$ is never greater than 1/2... So the improvement allows qualitative progress.

Both Theorem 2 and the Greendlinger property will be crucially used in [OW] to show that random groups at densities $< 1/6$ act freely cocompactly on $CAT(0)$ cube complexes and satisfy the Haagerup property.

Corollary 4 is probably unimportant but might justify to some extent the term "cancellation on average" applied to the density model (although this is certainly not "small cancellation on average", since when d is close to $1/2$ the cancellation becomes arbitrarily large).

The estimate of the hyperbolicity constant in Corollary 3 is of course not qualitatively different from the previous, 10^{20} times larger one.

Theorem 6 refers to the random presentation obtained by applying directly the definition of the density model. Note that in any δ-hyperbolic group, the set of words of length at most 8δ representing the identity constitutes a presentation of the group satisfying the Dehn algorithm ([Sho91], Theorem 2.12); however, this set of words is quite large, and computing it is feasible but tedious. Moreover this set of words does not in general satisfy the Greendlinger property, which is what is really needed in lots of applications.

What happens at $d = 1/5$ is not known (just as what happens for infiniteness or triviality at $d = 1/2$), but probably depends on more precise subexponential terms in the number of relators of the presentation, and so might not be very interesting.

Theorem 2 seems to remain valid in more general random group models when the lengths of the relators are not the same but lie within some bounded ratio (see [Oll04]). However I do not know if this is the case for Theorem 6.

Theorem 2 may also help show that random groups at different densities are indeed different.

Acknowledgements. Part of the ideas presented here arose during my stay in Montréal in July 2004 at the invitation of Daniel T. Wise, whom I would like to thank for helpful discussions and his so warm welcome. I would also like to thank Thomas Delzant who asked for the sharp isoperimetric constant in random groups, and insisted on the importance of the Dehn algorithm.

Local-global principles

Following Gromov ([Gro87], 2.3.F, 6.8.M), there have been lots of somewhat different phrasings of the local-global principle for hyperbolic groups (chapter 8 of [Bow91], [Ols91], [Bow95], [Pap96]). This principle states that to ensure hyperbolicity, it is enough to check the isoperimetric inequality on a *finite* number of diagrams.

We give here a version which can be very neatly applied in our context, and which involves arbitrarily small loss in the isoperimetric constant. Though this version is not difficult to prove using previously stated results, it does not seem to be a formal corollary thereof.

Definition 7.
Let D be a van Kampen diagram with respect to some presentation. The area $\mathcal{A}(D)$ *of D is the sum of the boundary lengths of all faces of D.*

We have advocated elsewhere ([Oll05], [Oll-a]) that this is the right way to measure area in a context of linear isoperimetric inequalities involving relators with very different lengths. That it allows a formulation of the local-global principle without loss in the constant is a further argument in this direction.

Theorem 8.
Let $G = \langle\, a_1, \ldots, a_m \mid R\,\rangle$ be a finite group presentation and let ℓ_1, ℓ_2 be the minimal and maximal lengths of a relator in R.

Let P be a class of van Kampen diagrams, such that any subdiagram of a diagram in P lies in P.

Let $C > 0$. Choose $\varepsilon > 0$. Suppose that for some $K \geqslant 10^{50}\,(\ell_2/\ell_1)^3\,\varepsilon^{-2}\,C^{-3}$, any van Kampen diagram D in P of area at most $K\ell_2$ satisfies

$$|\partial D| \geqslant C\,\mathcal{A}(D)$$

Then any van Kampen diagram D in P satisfies

$$|\partial D| \geqslant (C - \varepsilon)\,\mathcal{A}(D)$$

In particular, if P is such that for each reduced word w representing the identity in G, there is at least one diagram in P spanning w, then G is hyperbolic.

It is not clear whether ℓ_2/ℓ_1 really has an impact on the constants.

Typical useful examples of the class P are "reduced", or "of minimal area", or "of minimal number of faces" (minimal for a given boundary word).

This theorem may allow to extend the scope of the density model by taking relators of length between ℓ and $\ell^{1+\alpha}$ for some positive α, instead of taking relators of length exactly ℓ (see the discussions in [Oll04]).

We are going to state closer and closer propositions to the theorem. The first one is a variant on Papasoglu's exposition [Pap96] as modified in [Oll04].

Let X be a complex of dimension 2. 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 drawn in X* is a cellular map (maybe dimension-decreasing) from a cellular disk to X.

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.

We then have ([Oll04], Proposition 42, p. 666):

Proposition 9.
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 combinatorial area $A_c(D)$ 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 $A_c(D) \geqslant k^2$, satisfies

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

This allows to prove one more step:

Proposition 10.
Let $G = \langle\, a_1, \dots, a_m \mid R\,\rangle$ be a finite presentation and let ℓ_1, ℓ_2 be the minimal and maximal lengths of a relator in R.

Let P be a class of van Kampen diagrams, such that any subdiagram of a diagram in P lies in P.

Let $C > 0$. Suppose that for some $K \geqslant 10^{23}\,(\ell_2/\ell_1)\,C^{-2}$, any van Kampen diagram D in P of area $\mathcal{A}(D)$ at most $K\ell_2$ satisfies

$$|\partial D| \geqslant C\,\mathcal{A}(D)$$

Then any van Kampen diagram D in P satisfies

$$|\partial D| \geqslant C'\,\mathcal{A}(D)$$

with $C' = C\,(\ell_1/\ell_2)/10^{15}$.

Proof.
We have $A_c(D)/\ell_2 \leqslant \mathcal{A}(D) \leqslant A_c(D)/\ell_1$ for any diagram D in class P (remember $\mathcal{A}(D)$ is the sum of the lengths of the faces whereas $A_c(D)$ is the sum of the squares of these lengths).

Set $k^2 = K\ell_1\ell_2/480$. Let D be a van Kampen diagram such that $k^2/4 \leqslant A_c(D) \leqslant 480k^2$. We have $\mathcal{A}(D) \leqslant A_c(D)/\ell_1 \leqslant K\ell_2$. So the assumption of the proposition states that $L_c(D) = |\partial D| \geqslant C\mathcal{A}(D)$. Thus

$$L_c(D)^2 \geqslant C^2\mathcal{A}(D)^2 \geqslant C^2 A_c(D)^2/\ell_2^2 \geqslant C^2 A_c(D)k^2/4\ell_2^2 = A_c(D)C^2K(\ell_1/\ell_2)/1920$$

So if $k \geqslant 10^{10}\ell_2$ and $C^2K(\ell_1/\ell_2)/1920 \geqslant 2\cdot 10^{14}$ then the assumptions of Proposition 9 are fulfilled. Taking $K = 10^{23}\,(\ell_2/\ell_1)/C^2$ is enough to ensure this is the case. The consequence of Proposition 9 is then that

$$|\partial D| = L_c(D) \geqslant A_c(D)/10^4k \geqslant \mathcal{A}(D)\,\ell_1/10^4k$$

and unwinding the constants shows that $\ell_1/10^4k \geqslant C\,(\ell_1/\ell_2)/10^{15}$. □

Going on with our approximations of Theorem 8, we now know that there exists an isoperimetric constant C', but its value may be much smaller than the original constant C. We solve the problem by a kind of bootstrapping: we will re-do some kind of local-global passage, using our knowledge of hyperbolicity of the group. This will allow to keep the constants tight.

We need a lemma from [Oll05].

The *distance to boundary* of a face of a van Kampen diagram is the minimal length of a sequence of faces adjacent by an edge, beginning with the given face and ending with a face adjacent to the boundary (so that a boundary face is at distance 1 from the boundary).

Let C' be the isoperimetric constant provided by Proposition 10, so that any diagram D in P satisfies $|\partial D| \geqslant C'\mathcal{A}(D)$. Set

$$\alpha = 1/\log(1/(1-C')) \leqslant 1/C'$$

The following is Lemma 10 of [Oll05], where we replaced "minimal" by "in class P".

Lemma 11.
Let D be a van Kampen diagram in class P. Then D can be partitioned into two diagrams D', D'' by cutting it along a path of length at most $\ell_2 + 2\alpha\ell_2\log(\mathcal{A}(D)/\ell_2)$ with endpoints on the boundary of D, such that each of D' and D'' contains at least one quarter of the boundary of D.

With this we can get closer to Theorem 8.

Proposition 12.
Let $G = \langle\, a_1,\dots,a_m \mid R\,\rangle$ be a finite presentation and let ℓ_1,ℓ_2 be the minimal and maximal lengths of a relator in R.

Let P be a class of van Kampen diagrams, such that any subdiagram of a diagram in P lies in P.

Let $C, C' > 0$. Choose some $\varepsilon > 0$. Suppose that any van Kampen diagram D in P satisfies

$$|\partial D| \geqslant C'\mathcal{A}(D)$$

and that, for some $A \geqslant 50/(\varepsilon C')^2$, any van Kampen diagram D in P having boundary length at most $A\ell_2$ satisfies

$$|\partial D| \geqslant C\mathcal{A}(D)$$

Then any van Kampen diagram D in P, with boundary length at most $7A\ell_2/6$, satisfies

$$|\partial D| \geqslant (C-\varepsilon)\,\mathcal{A}(D)$$

Proof.
Let D be a van Kampen diagram in P, of boundary length between $A\ell_2$ and $7A\ell_2/6$. By the isoperimetry assumption for all diagrams we have $\mathcal{A}(D) \leqslant 7A\ell_2/6C'$.

By Lemma 11, we can partition D into two diagrams D' and D'', each of them containing at least one quarter of the boundary length of D. So we have $|\partial D'| \leqslant 3\,|\partial D|\,/4 + \ell_2(1 + 2\alpha \log(7A/6C')) \leqslant \ell_2\,(7A/8 + 1 + 2\alpha \log(7A/6C'))$ and likewise for D''.

Choose A large enough (depending only on C') so that $1 + 2\alpha \log(7A/6C') \leqslant A/8$. Then both D' and D'' have boundary length at most $A\ell_2$. So by assumption we have

$$\left|\partial D'\right| \geqslant C\mathcal{A}(D') \text{ and } \left|\partial D''\right| \geqslant C\mathcal{A}(D'')$$

(note the occurrence of C and not C').

Now we choose A large enough (again depending only on C') so that $2{+}4\alpha \log(7A/6C') \leqslant \varepsilon A$ (if we remember that $\alpha \leqslant 1/C'$, taking $A = 50/(\varepsilon C')^2$ is enough). We have

$$\begin{aligned}
|\partial D| &= \left|\partial D'\right| + \left|\partial D''\right| - 2\left|\partial D' \cap \partial D''\right| \\
&\geqslant \left|\partial D'\right| + \left|\partial D''\right| - \ell_2\left(2 + 4\alpha \log(7A/6C')\right) \\
&\geqslant C\left(\mathcal{A}(D') + \mathcal{A}(D'')\right) - \varepsilon A\ell_2 \\
&\geqslant (C - \varepsilon)\,\mathcal{A}(D)
\end{aligned}$$

since $\mathcal{A}(D') + \mathcal{A}(D'') = \mathcal{A}(D) \geqslant |\partial D| \geqslant A\ell_2$. □

The last approximation to Theorem 8 is the following:

Proposition 13.
Let $G = \langle\, a_1, \dots, a_m \mid R\,\rangle$ be a finite presentation and let ℓ_1, ℓ_2 be the minimal and maximal lengths of a relator in R.

Let P be a class of van Kampen diagrams, such that any subdiagram of a diagram in P lies in P.

Let $C, C' > 0$. Choose some $\varepsilon > 0$. Suppose that any van Kampen diagram D in P satisfies

$$|\partial D| \geqslant C'\mathcal{A}(D)$$

and that, for some $K \geqslant 50/(\varepsilon^2 {C'}^3)$, any van Kampen diagram D in P having area at most $K\ell_2$ satisfies

$$|\partial D| \geqslant C\mathcal{A}(D)$$

Then any van Kampen diagram D in P satisfies

$$|\partial D| \geqslant (C - 14\varepsilon)\,\mathcal{A}(D)$$

Proof.
Set $A = C'K$. Let D be a diagram in P of boundary length at most $A\ell_2$. By the assumption on all diagrams, D has area at most $A\ell_2/C' = K\ell_2$ so that by the assumption on small diagrams we have $|\partial D| \geqslant C\,\mathcal{A}(D)$. In particular, the assumptions of Proposition 12 are fulfilled.

So this proposition implies that diagrams D in P of area at most $7A\ell_2/6$ satisfy $|\partial D| \geqslant (C-\varepsilon)\,\mathcal{A}(D)$. This means that the assumptions of Proposition 12 are fulfilled with the new parameters $A_1 = 7A/6$, $\varepsilon_1 = \varepsilon(6/7)^{1/2}$ and $C_1 = C-\varepsilon$ instead of A, ε, C, and with the same C' (these new parameters indeed satisfy $A_1 \geqslant 50/(\varepsilon_1 C')^2$).

So applying Proposition 12 again, we get that diagrams D in P of area at most $A_2 = A\ell_2(7/6)^2$ satisfy $|\partial D| \geqslant C_2\,\mathcal{A}(D)$ where $C_2 = C_1 - \varepsilon_1$.

By induction, we get that diagrams D in P of area at most $A\ell_2(7/6)^k$ satisfy

$$|\partial D| \geqslant \left(C - \varepsilon \sum_{i=0}^{k-1} (6/7)^{i/2}\right) \mathcal{A}(D)$$

and we conclude by the inequality $\sum_{i=0}^{\infty}(6/7)^{i/2} < 14$. □

Proof of Theorem 8.
Applying Proposition 10 (which is allowed since $10^{50}\,(\ell_2/\ell_1)^3\,\varepsilon^{-2}\,C^{-3} \geqslant 10^{23}\,(\ell_2/\ell_1)\,C^{-2}$), we get that any van Kampen diagram D in P satisfies $|\partial D| \geqslant C'\,\mathcal{A}(D)$ where $C' = C\,(\ell_1/\ell_2)/10^{15}$. We conclude with Proposition 13 (where we replace ε with $\varepsilon/14$). □

Proof of Theorem 2

Now Theorem 2 is an easy consequence of Theorem 8 and already known facts about random groups. First, we recall the result from [Gro93] (see also [Oll04]) on diagrams of bounded size.

Suppose we are given a random presentation at density d, by reduced relators of length ℓ.

Theorem 14 (M. Gromov).
For every $\varepsilon > 0$ and every $K \in \mathbb{N}$, with overwhelming probability, every reduced van Kampen diagram with at most K faces satisfies

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

Of course, the point is that the overwhelming probability is a priori not uniform in K.

Proof.
We only have to change a little bit the conclusion of the proof in [Oll04], p. 613. It is proven there that if D is a reduced van Kampen diagram involving $n \leqslant |D|$ distinct relators $r_1, \dots, r_n$, with relator r_i appearing m_i times in the diagram (we can assume $m_1 \geqslant \dots \geqslant m_n$), then there exist numbers d_i, $1 \leqslant i \leqslant n$ such that:

$$|\partial D| \geqslant (1-2d)\,\ell\,|D| + 2\sum d_i(m_i - m_{i+1})$$

and such that the probability of this situation is at most $(2m)^{\inf d_i}$ ([Oll04], p. 613). In particular, for fixed ε, with overwhelming probability we can suppose that $\inf d_i \geqslant -\ell\varepsilon/2$.

If all d_i's are non-negative, then we get $|\partial D| \geqslant (1-2d)\,\ell\,|D|$ as needed.

Otherwise, as $1 \leqslant m_i \leqslant |D|$ and $m_i \geqslant m_{i+1}$ we have $\sum d_i(m_i - m_{i+1}) \geqslant |D| \inf d_i$ and so

$$|\partial D| \geqslant (1-2d)\,\ell\,|D| + 2\,|D| \inf d_i \geqslant (1-2d-\varepsilon)\,\ell\,|D|$$

□

Proof of Theorem 2.
Theorem 2 now is an immediate consequence of Theorem 14 and Theorem 8 (where the class P is the class of all reduced diagrams). □

Proof of Corollary 3.
For Corollary 3 we use the following proposition, which is only a weaker version, adapted to our vocabulary, of Lemma 3.11 of [Cha94]:

Proposition 15.
Suppose that a finite group presentation satisfies the following: for every reduced word w representing the identity in the group, there exists a van Kampen diagram D spanning w with $|\partial D| \geqslant C\mathcal{A}(D)$. Let λ be the maximal length of a relator in the presentation.

Then the group is δ-hyperbolic with $\delta < 4\lambda/C$ (w.r.t. the metric defined by the generators in the presentation).

Indeed, Lemma 3.11 of [Cha94] states that for some notion of area $\mathrm{area}_{\mathrm{Champetier}}$, the isoperimetric inequality $\mathrm{area}_{\mathrm{Champetier}}(D) \leqslant \alpha\,|\partial D|$ for van Kampen diagrams (actually for curves in a geodesic metric space) implies δ-hyperbolicity with $\delta \leqslant 20\alpha$.

The notion of area used by Champetier (Definition 3.2 in [Cha94]) is different from $\mathcal{A}(D)$ as defined in this paper. However it is noted by Champetier that a curve of length L has $\mathrm{area}_{\mathrm{Champetier}} \leqslant L^2/2\pi$. So for a van Kampen diagram D we have

$$\mathrm{area}_{\mathrm{Champetier}}(D) \leqslant \sum_{f \text{ face of } D} \frac{|\partial f|^2}{2\pi} \leqslant \frac{\lambda}{2\pi} \sum_{f \text{ face of } D} |\partial f| = \frac{\lambda\mathcal{A}(D)}{2\pi}$$

Consequently the inequality $|\partial D| \geqslant C\mathcal{A}(D)$ implies that the Champetier assumption $\mathrm{area}_{\mathrm{Champetier}}(D) \leqslant \alpha\,|\partial D|$ holds with $\alpha = \lambda/(2C\pi)$, hence Proposition 15 noting that $20/2\pi < 4$.

Corollary 3 now follows from Theorem 2 and Proposition 15, choosing ε small enough. □

Proof of Corollary 4.
Corollary 4 is easy. Let $D = D_1 \cup_w D_2$. Since $|\partial D| \geqslant (1-2d-\varepsilon)\,\ell\,|D|$, the number of

internal edges of D is at most $(d+\varepsilon/2)\ell\,|D|$. So a fortiori $|w| \leqslant (d+\varepsilon/2)\ell\,|D|$. Now

$$\begin{aligned}|w| \leqslant (d+\varepsilon/2)\,\ell\,|D| &\leqslant \frac{d+\varepsilon}{1-2d-\varepsilon}\,|\partial D| \\ &= \frac{d+\varepsilon/2}{1-2d-\varepsilon}\,(|\partial D_1|+|\partial D_2|-2\,|w|)\end{aligned}$$

and so

$$|w| \leqslant (d+\varepsilon/2)\;(|\partial D_1|+|\partial D_2|)$$

as needed. □

Dehn's algorithm and Greendlinger's Property

We now turn to the proof of Theorem 6. Since the Greendlinger property is stronger than the Dehn algorithm, it suffices to prove the former for $d<1/5$ and disprove the latter for $d>1/5$.

Greendlinger's Property for $d<1/5$. We begin by a lemma which is weaker in the sense that we do not ask for the boundary edges to be consecutive. We will then conclude by a standard argument.

Lemma 16.
For any $\varepsilon>0$, with overwhelming probability, at density d the following holds:

Let D be a reduced van Kampen diagram with at least two faces. There exist two faces of D each having at least $\ell(1-5d/2-\varepsilon)$ edges on the boundary of D (maybe not consecutive).

Observe that when $d<1/5$ this is more than $\ell/2$ (for small enough ε depending on $1/5-d$). This lemma is also valid at densities larger than $1/5$ but becomes trivial at $d=2/5$.

Proof of the lemma.
Let D be a reduced van Kampen diagram with at least two faces.

First, note that it is enough to consider the case when D is homeomorphic to a disk. Otherwise, decompose D as the union of "filaments" and maximal parts homeomorphic to a disk. Adding or removing filaments does not change the property of a face having so many edges on the boundary of the diagram.

Let f be a face of D having the greatest number of edges on the boundary. Say f has $\alpha\ell$ edges on the boundary. Suppose that any face other than f has no more than $\beta\ell$ edges on the boundary. We want to show that $\beta \geqslant 1-5d/2-\varepsilon$. So suppose that $\beta<1-5d/2-\varepsilon$. (The reader may find more convenient to read the following skipping the ε's.)

Consider also the (maybe not connected, but this does not matter) diagram D' obtained by removing face f from D. We have $|\partial D'| = |\partial D|+\ell-2\alpha\ell$.

By definition of α and β, and since D is homeomorphic to a disk, we have $|\partial D| \leqslant \beta\ell(|D|-1)+\alpha\ell$. Consequently $|\partial D'| \leqslant \beta\ell(|D|-1)+\ell-\alpha\ell$.

But by Theorem 2, with overwhelming probability we can suppose that we have $|\partial D| \geqslant (1-2d-\varepsilon/2)\,\ell\,|D|$ and $|\partial D'| \geqslant (1-2d-\varepsilon/2)\,\ell\,|D'| = (1-2d-\varepsilon/2)\,\ell\,(|D|-1)$. So combining these inequalities we get

$$\begin{gathered}(1-2d-\varepsilon/2)\,|D| \leqslant \beta(|D|-1)+\alpha\\ (1-2d-\varepsilon/2)\,(|D|-1) \leqslant \beta(|D|-1)+1-\alpha\end{gathered}$$

or, since we assumed by contradiction that $\beta < 1-5d/2-\varepsilon$,

$$\begin{gathered}(1-2d-\varepsilon/2)\,|D| < (1-5d/2-\varepsilon)\,(|D|-1)+\alpha\\ (1-2d-\varepsilon/2)\,(|D|-1) < (1-5d/2-\varepsilon)\,(|D|-1)+1-\alpha\end{gathered}$$

which yield respectively

$$|D| < \frac{\alpha+5d/2-1+\varepsilon}{d/2+\varepsilon/2} \tag{1}$$

$$|D| < \frac{d/2+1-\alpha+\varepsilon/2}{d/2+\varepsilon/2} \tag{2}$$

Either $\alpha \leqslant 1-d-\varepsilon/4$ or $\alpha \geqslant 1-d-\varepsilon/4$. In any case, one of (1) or (2) gives

$$|D| < \frac{3d/2+3\varepsilon/4}{d/2+\varepsilon/2} < 3$$

(generally, a face having more than $(1-d)\ell$ on the boundary is the frontier at which it is more interesting to remove this face before applying Theorem 2).

The case $|D| \leqslant 2$ is easily treated by Theorem 2. So we get a contradiction, and the lemma is proven. □

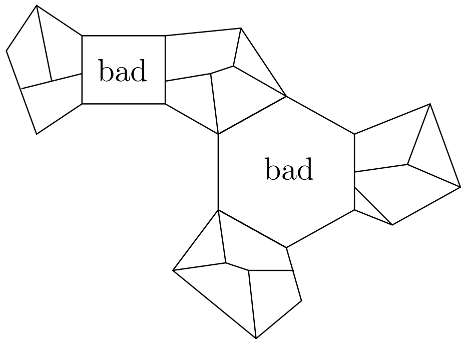
This somewhat obscure proof and the role of 1/5 will become clearer in the next paragraph, when we will build a 3-face diagram for $d > 1/5$ with only one face having more than $\ell/2$ boundary edges.

Back to the proof of Greendlinger's Property for $d < 1/5$. If we are facing a diagram D such that the intersection of the boundary of any face of D with the boundary of D is connected, then Lemma 16 provides what we want.

Now we apply a standard argument to prove that this case is enough. Suppose that some face of D has a non-connected intersection with the boundary, having two (or more) boundary components, so that this face separates the rest of the diagram into two (or more) components. Call *good* a face having exactly one boundary component and *bad* a face with two or more boundary components (there are also internal faces, which we are not interested in).

First suppose that D is homeomorphic to a disk (so that no single edge or vertex removal can disconnect it).

Decompose D into bad faces and maximal parts without bad faces. Call such a maximal part *extremal* if it is in contact with only one bad face. It is clear that, if there exists some bad face, there are at least two such extremal parts.



To reach the conclusion it is sufficient to find in any extremal part a good face having more than $\ell(1 - 5d/2 - \varepsilon)$ edges on the boundary. So let f be a bad face in contact with an extremal part P without bad faces.

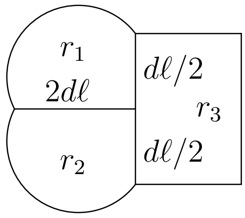
Consider the diagram $D' = P \cup f$. This diagram has no bad faces now, and so by Lemma 16 there are two faces in it having more than $\ell(1 - 5d/2 - \varepsilon)$ consecutive edges on the boundary. One of these may be f, but the other one has to be in P and so has more than $\ell(1 - 5d/2 - \varepsilon)$ consecutive edges on the boundary of D as well.

Now in the case D is not homeomorphic to a disk, then the "filaments" (the edges/vertices the removal of which disconnects D) are treated the same way as bad faces in the previous argument.

A counter-example for $d > 1/5$. Here we show that the presentation does not satisfy the Dehn algorithm as soon as $d > 1/5$.

Fix some $\varepsilon > 0$. We can with overwhelming probability find two relators r_1, r_2 sharing a common subword w of length $(2d - \varepsilon)\ell$. Once those are chosen, let x be the subword of length $(d - \varepsilon)\ell$ of the boundary of the diagram $r_1 \cup_w r_2$ occurring around some endpoint of the w-gluing and having length $(d - \varepsilon)\ell/2$ on each side of this endpoint (see picture below). (When $d > 2/5$ there is less than this left on the boundary of $r_1 \cup_w r_2$; but the situation is even easier at larger densities and so we leave this detail aside).

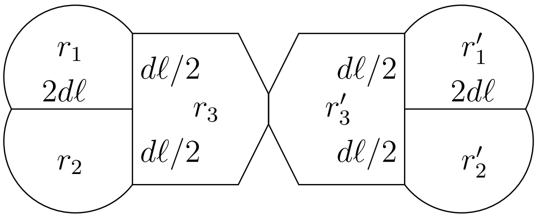
At density d, subwords of length $(d-\varepsilon)\ell$ of the relators exhaust all reduced words of length $(d-\varepsilon)\ell$. So it is possible to find a relator r_3 gluing to $r_1 \cup_w r_2$ along x. After this operation r_1 and r_2 each have less than $1-(2d-\varepsilon)\ell-(d/2-\varepsilon/2)\ell = (1-5d/2+3\varepsilon/2)\ell$ of their length on the boundary of the new diagram (see picture below), which is less than $\ell/2$ when $d > 1/5$, for small enough ε. Compare Lemma 16 — which is thus sharp.



This diagram violates the Greendlinger property (but not yet the Dehn algorithm). Note for later use that at this step, the boundary length of the diagram so obtained is $(3-6d+4\varepsilon)\ell$. This is the smallest possible value compatible with Theorem 2, up to the ε's.

But (thanks to the ε's) this will not only happen once but arbitrarily many times as $\ell \to \infty$, so we can find another independent triple of relators (r'_1, r'_2, r'_3) giving rise to the same configuration.

Now if r_3 and r'_3 share only a single letter in the region of length $\ell/5$ opposite to the position where they glue to $r_1 \cup_w r_2$ (resp. $r'_1 \cup_{w'} r'_2$) (and this happens all the time thanks to the law of large numbers), then we can form a diagram in which r_3 and r'_3 become faces having no more than $\ell/2$ consecutive edges on the boundary (they are bad faces in the terminology of the previous proof). So if $d > 1/5$, no face of this diagram has more than $\ell/2$ consecutive edges on the boundary (although the two bad faces have more than $\ell/2$ non-consecutive boundary edges).



This is not enough to disprove the Dehn algorithm: this algorithm only demands that for any reduced word representing e, there exists *some* van Kampen diagram with the boundary face property. There could exist another van Kampen diagram D' with the same boundary word as the diagram D above, in which some face would have more than $\ell/2$ consecutive edges on the boundary. So let r_4 be this face. Since D and D' have the same boundary word, we can glue r_4^{-1} to the previous diagram D to get a new diagram D'' with 7 faces; since r_4 has more than half of its length on the boundary of D' we have $|\partial D''| < |\partial D|$.

Either D' is reduced or r_4 is equal to some relator r_i already present in the diagram.

In the former case, we get that $|\partial D| = (3-6d+4\varepsilon)\ell \times 2 - 2 = 6(1-2d)\,\ell + 8\varepsilon\ell - 2$. Since $|\partial D'| < |\partial D|$ we get $|\partial D'| < 6(1-2d)\,\ell + 8\varepsilon\ell - 2$. But by Theorem 2, for any ε' we have $|\partial D'| \geqslant 7(1-2d-\varepsilon')\ell$, which is a contradiction for small enough values of ε and ε'.

In the latter case, this means that we can glue a copy of r_i^{-1} along r_i on the boundary of the diagram D along more than $\ell/2$ edges. But, since the r_i included in D has no more than $\ell/2$ consecutive edges on the boundary, this means that before gluing r_i^{-1} we could have folded some letters of r_i with neighbouring letters in the boundary of D. This is excluded if we assume (as we can always do) that the boundary of D is reduced.

References

[Bow91] B. H. Bowditch, *Notes on Gromov's hyperbolicity criterion for path-metric spaces*, in *Group theory from a geometrical viewpoint*, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991), 64–167.

[Bow95] B. H. Bowditch, *A short proof that a subquadratic isoperimetric inequality implies a linear one*, Michigan Math. J. **42** (1995), n° 1, 103–107.

[Cha94] Ch. Champetier, *Petite simplification dans les groupes hyperboliques*, Ann. Fac. Sci. Toulouse Math., ser. 6, vol. 3 (1994), n°2, 161–221.

[Ghy03] É. Ghys, *Groupes aléatoires*, séminaire Bourbaki **916** (2003).

[Gre60] M. Greendlinger, *Dehn's algorithm for the word problem*, Comm. Pure Appl. Math. **13** (1960), 67–83.

[Gro87] M. Gromov, *Hyperbolic Groups*, in *Essays in group theory*, ed. S.M. Gersten, Springer (1987), 75–265.

[Gro93] M. Gromov, *Asymptotic Invariants of Infinite Groups*, in *Geometric group theory*, ed. G. Niblo, M. Roller, Cambridge University Press, Cambridge (1993).

[Gro03] M. Gromov, *Random Walk in Random Groups*, Geom. Funct. Anal. **13** (2003), n° 1, 73–146.

[LS77] R.C. Lyndon, P.E. Schupp, *Combinatorial Group Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete **89**, Springer (1977).

[Oll04] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, GAFA, Geom. Funct. Anal. **14** (2004), n° 3, 595–679.

[Oll05] Y. Ollivier, *Cogrowth and spectral gap of generic groups*, Ann. Inst. Fourier (Grenoble) **55** (2005), n° 1, 289–317.

[Oll-a] Y. Ollivier, *Growth exponent of generic groups*, preprint (2003), ArXiV `math.GR/0401050`

[Oll-b] Y. Ollivier, *A January 2005 invitation to random groups*, expository manuscript (2005).

[Ols91] A.Yu. Ol'shanksiĭ, *Hyperbolicity of groups with subquadratic isoperimetric inequality*, Int. J. Algebra Comput. **1** (1991), n° 3, 281–289.

[OW] Y. Ollivier, D. T. Wise, *Cubulating groups at density* $< 1/6$, preprint (2005).

[Pap96] 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.

[Sho91] H. Short et al., *Notes on word hyperbolic groups*, in *Group Theory from a Geometrical Viewpoint*, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991), 3–63.

Habilitation à diriger des recherches

Cubulating random groups at density less than 1/6

Ce texte est écrit en collaboration avec Dani Wise, de l'université McGill à Montréal. Après notre rencontre au Japon, Dani m'a invité à Montréal en août 2004, séjour qui a marqué le début de ce travail. Ce texte est accepté pour publication dans les Transactions of the American Mathematical Society.

Cubulating random groups at density less than 1/6

Yann Ollivier & Daniel T. Wise

Abstract

We prove that random groups at density less than $\frac{1}{6}$ act freely and cocompactly on CAT(0) cube complexes, and that random groups at density less than $\frac{1}{5}$ have codimension-1 subgroups. In particular, Property (T) fails to hold at density less than $\frac{1}{5}$.

Introduction

Gromov introduced in [Gro93] the notion of a random finitely presented group on $m \geqslant 2$ generators at density $d \in (0;1)$. The idea is to fix a set $\{g_1, \ldots, g_m\}$ of generators and to consider presentations with $(2m-1)^{d\ell}$ relations each of which is a random reduced word of length ℓ (Definition 1). The *density* d is a measure of the size of the number of relations as compared to the total number of available relations. See Section 1 for precise definitions and basic properties, and [Oll05b, Gro93, Ghy04, Oll04] for a general discussion on random groups and the density model.

One of the striking facts Gromov proved is that a random finitely presented group is infinite, hyperbolic at density $< \frac{1}{2}$, and is trivial or $\{\pm 1\}$ at density $> \frac{1}{2}$, with probability tending to 1 as $\ell \to \infty$.

Żuk obtained Property (T) for a related class of presentations at density $> \frac{1}{3}$ (see [Żuk03] and the discussion in [Oll05b]). On the other hand, Gromov observed that at density $< d$, a random presentation satisfies the $C'(2d)$ small cancellation condition. Consequently, at density $< \frac{1}{12}$, the groups will not have Property (T) since $C'(\frac{1}{6})$ groups act properly discontinuously on CAT(0) cube complexes [Wis04].

As above, the statements about the behavior of a group at a certain density are only correct with probability tending to 1 as $\ell \to \infty$. Throughout the paper, we will say that a given property holds *with overwhelming probability* if its probability tends exponentially to 1 as $\ell \to \infty$.

The goals of this paper are a complete geometrization theorem at $d < \frac{1}{6}$, implying the Haagerup property, and existence of a codimension-1 subgroup at $d < \frac{1}{5}$, implying failure of Property (T):

Theorem 62.
With overwhelming probability, random groups at density $d < \frac{1}{6}$ act freely and cocompactly on a CAT(0) *cube complex.*

Corollary 56.
With overwhelming probability, random groups at density $d < \frac{1}{6}$ are a-T-menable (Haagerup property).

Theorem 50.
With overwhelming probability, random groups G at density $d < \frac{1}{5}$ have a subgroup H which is free, quasiconvex and such that the relative number of ends $e(G, H)$ is at least 2.

Corollary 51.
With overwhelming probability, random groups at density $d < \frac{1}{5}$ do not have Property (T).

CAT(0) cube complexes are a higher dimensional generalization of trees, which arise naturally in the splitting theory of groups with codimension-1 subgroups [Sag95, Sag97]. A group is *a-T-menable* or has the *Haagerup property* [CCJ+01] if it admits a proper isometric action on a Hilbert space. This property is, in a certain sense, an opposite to Kazhdan's *Property* (T) [dlHV89, BdlHV08] which (for second countable, locally compact groups) is characterized by the requirement that every isometric action on an affine Hilbert space has a fixed point. There is also a definition of the Haagerup property in terms of a proper action on a space with measured walls [CMV04, CDH], which is a natural framework for some of our results.

The *relative number of ends* $e(G, H)$ of the subgroup H of the finitely generated group G is the number of ends of the Schreier coset graph $H\backslash G$ (see [Hou74, Sco78]). Note that $e(G, H)$ is independent of the choice of a finite generating set. We say H is a *codimension*-1 *subgroup* of G if H coarsely disconnects the Cayley graph Γ of G, in the sense that the complement $\Gamma - N_k(H)$ of some neighborhood of H contains at least two components that are not contained in any finite neighborhood $N_j(H)$ of H. The above two notions are very closely related and are sometimes confused in the literature: If $e(G, H) > 1$ then H is a codimension-1 subgroup of G, and the converse holds when there is more than one H-orbit of an "infinitely deep" component in $\Gamma - N_k(H)$.

Let us present the structure of the argument. In [Sag95], Sageev gave a fundamental construction which, from a codimension-1 subgroup H of G, produces an "essential" action of G on a CAT(0) cube complex. From [NR97] or [NR98] we know, in turn, that groups acting essentially/properly on a CAT(0) cube complex, act essentially/properly on a Hilbert space and cannot have Property (T) (their proof is a generalization of a proof in [BJS88] that infinite Coxeter groups are a-T-menable, which in turn, was a generalization of Serre's argument that an essential action on a tree determines an essential action on a Hilbert space [Ser80]).

In our situation the codimension-1 subgroups will arise as stabilizers of some codimension-1 subspaces, called *hypergraphs*, in the Cayley 2-complex $\widetilde{X}$ of the random group G. These hypergraphs are the same as those in [Wis04] and are defined in Section 2. The basic idea is, from the midpoint of each 1-cell in a 2-cell c, to draw a line to the midpoint of the opposite 1-edge in c (assuming all 2-cells have even boundary length). These lines draw a graph in the 2-complex, whose connected components are the hypergraphs. Hypergraphs are natural candidates to be walls [HP98].

In Section 4 we show that at density $d < \frac{1}{5}$, with overwhelming probability, the hypergraphs embed (quasi-isometrically) in the Cayley 2-complex. The main idea is that if a hypergraph self-intersects, it will circle around a disc in the Cayley 2-complex, thus producing a *collared diagram* (Section 3). But at $d < \frac{1}{5}$, the Dehn algorithm holds for a random group presentation [Oll07], so that in each van Kampen diagram some 2-cell has more than half its length on the boundary, which is impossible if a hypergraph runs around the boundary 2-cells of the diagram.

A consequence of this embedding property is that each hypergraph is a tree dividing $\widetilde{X}$ into two connected components, thus turning $\widetilde{X}$ into a space with walls [HP98].

We then show that these walls can be used to define (free quasiconvex) codimension-1 subgroups (Section 7). For this we need the complex $\tilde{X}$ to go "infinitely far away" on the two sides of a given wall. This is guaranteed by exhibiting a pair of infinite hypergraphs intersecting at only one point. At $d < \frac{1}{6}$, hypergraphs intersect at at most one point except for a degenerate case (Section 5). This is not true in general for $d < \frac{1}{5}$; however, one can still prove that through a "typical" 2-cell that a hypergraph Λ_1 passes through, there passes a second hypergraph Λ_2 transverse to Λ_1, which is enough (Section 6).

To prove the Haagerup property, we show that at $d < \frac{1}{6}$, the number of hypergraphs separating given points $p, q \in \tilde{X}$ is at least $\mathrm{dist}_{\tilde{X}}(p, q)/K$ for some constant K. Consequently the wall metric is quasi-isometric to the Cayley graph metric, which implies that the group has the Haagerup property. Key objects here are *hypergraph carriers* (the set of 2-cells through which a hypergraph travels): at $d < \frac{1}{6}$ these carriers are convex subcomplexes of $\tilde{X}$, but this is not the case at $d > \frac{1}{6}$. We were unable to prove the separation by hypergraphs property at $d < \frac{1}{5}$ where the failure of convexity substantially complicates matters, though we conjecture such a statement still holds.

Finally, Theorem 62 is proven by combining the various properties established at $d < \frac{1}{6}$ (including, most importantly, the separation of points by a linear number of hypergraphs) to see that the cubulation criteria in [HW04] are satisfied; these criteria guarantee that the action of G on the CAT(0) cube complex associated with a codimension-1 subgroup arising from hypergraphs is indeed free and cocompact.

At density $d > \frac{1}{5}$, our approachs completely fails: with overwhelming probability, there is only one hypergraph Λ, which passes through every 1-cell of the Cayley complex (Section 11). Its stabilizer is the entire group, and it is thus certainly not codimension-1. We do not know if there are codimension-1 subgroups at density $\frac{1}{5} < d < \frac{1}{3}$. But, as mentioned above, the transition at $d = \frac{1}{5}$ in the behavior of hypergraphs is related to another one, namely failure of the Dehn algorithm for $d > \frac{1}{5}$ [Oll07], and our intuition is that something of both combinatorial and geometric relevance really happens at $d = \frac{1}{5}$.

1 Preliminaries and facts regarding Gromov's density

The density model of random groups was introduced by Gromov in [Gro93], Chapter 9 as a way to study properties of "typical" groups depending on the quantity of relators in a presentation of the group. We refer to [Oll05b, Gro93, Ghy04, Oll04] for general discussions on random groups and the density model.

Definition 1 (Density model of random groups).
Let $m \geqslant 2$ be an integer and consider the free group F_m generated by $a_1^{\pm 1}, \dots, a_m^{\pm 1}$.

Let $0 \leqslant d \leqslant 1$ be a density parameter. Let ℓ be a (large) length. Choose $(2m-1)^{d\ell}$ times (rounded to the nearest integer) at random a reduced word of length ℓ in the letters $a_1^{\pm 1}, \dots, a_m^{\pm 1}$, uniformly among all such words. Let R be the set of words so obtained.

A random group at density d and length ℓ *is the group $G = F_m/\langle R\rangle$, whose presentation is $\langle a_1, \dots, a_m \mid R \rangle$.*

A property is said to occur with overwhelming probability *in this model, if its probability of occurrence tends exponentially to* 1 *as $\ell \to \infty$.*

The basic intuition is that at density d, subwords of length $(d-\varepsilon)\ell$ of the relators in the presentation will exhaust all reduced words of length $(d-\varepsilon)\ell$.

The interest of the model is established through the following sharp phase transition theorem, proven by Gromov [Gro93] (see also [Oll04]):

Theorem 2 (M. Gromov).
Let $d < 1/2$. Then with overwhelming probability, a random group at density d is infinite, hyperbolic, torsion-free.

Let $d > 1/2$. Then with overwhelming probability, a random group at density d is either $\{1\}$ or $\{1, -1\}$.

One of the motivations for the results in this paper is the following ([Żuk03], see also the discussion in [Oll05b]):

Theorem 3 (A. Żuk).
Let $d > 1/3$. Then with overwhelming probability, a random group at density d has Property (T).

It is not known whether 1/3 is optimal in this theorem. Our results imply that 1/5 is a lower bound.

Remark 4.
According to the definition above, all relators in a random group have exactly the same length. However, the results stay the same if we take relators of length between ℓ and $\ell + C$ where C is any constant independent of ℓ.

Some results on random groups, including Theorem 2 and Theorem 3, also extend to the case when relators are taken of length between ℓ and $C\ell$ for some $C > 1$ (see [Oll04]), but we do not know if this is the case for the main theorems presented in this paper.

Hyperbolicity of random groups at $d < 1/2$ is proven using isoperimetry of van Kampen diagrams. In this paper we shall repeatedly need a precise statement of this isoperimetric inequality, which we state now.

For a van Kampen diagram D, we use the notation $|\partial D|$ for the length of its boundary path, and the notation $|D|$ for the number of 2-cells in D.

Convention 5.
When a property of a random group depends on a parameter ε, the phrase "the property occurs with overwhelming probability" will mean that for any $\varepsilon > 0$, the probability of the property tends to 1 as $\ell \to \infty$. This may not be uniform in ε.

The following, proven in [Oll07], is a strengthening of the original statement of Gromov, which held only for diagrams of size bounded by some constant. Note the role of $d = \frac{1}{2}$.

Theorem 6.
At density d, for any $\varepsilon > 0$ the following property occurs with overwhelming probability: all reduced van Kampen diagrams D satisfy

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

When using this result in this paper we will often omit the ε.

We now gather some definitions pertaining to small cancellation. We refer to chapter V of [LS77] for the definition of a *piece* in a group presentation.

Definition 7 (Small cancellation).
A presentation satisfies the $C'(\alpha)$ condition, *with* $0 \leqslant \alpha \leqslant 1$, *if for each relator R, and each piece P occurring in R, we have $|P| < \alpha\,|R|$.*

A presentation satisfies the $B(2p)$ condition *if no word that is more than half of a relator is the concatenation of p pieces.*

A presentation satisfies the $C(p)$ condition *if no relator R is the concatenation of fewer than p pieces.*

Note that $C'(\frac{1}{2p}) \Rightarrow B(2p) \Rightarrow C(2p)$ but that none of the reverse implications hold.

Proposition 8.
With overwhelming probability:

1. *The $C'(\alpha)$ condition occurs at density $< \alpha/2$.*
2. *The $B(6)$ condition occurs at density $< \frac{1}{8}$.*
3. *The $C(p)$ condition occurs at density $< \frac{1}{p}$.*

Proof.
The proof for $C'(\alpha)$ is written in detail in [Gro93], § 9.B. Let us briefly recall the

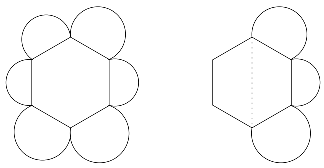


Figure 1: Diagrams contradicting the $C(7)$ and $B(6)$ conditions.

argument. Since the number of reduced words of length L is $(2m)(2m-1)^{L-1}$, the probability that two random reduced words of length ℓ share a common initial subword of length $L \leqslant \ell$ is $\left((2m)(2m-1)^{L-1}\right)^{-1} \leqslant (2m-1)^{-L}$.

So given two random words of length ℓ, the probability that they share a piece of length L is less than $\ell^2(2m-1)^{-L}$ where the ℓ^2 accounts for the choice of the position at which the piece occurs.

Now in a random group at density d, there are by definition $(2m-1)^{d\ell}$ relators. So the probability that there exists a couple of relators in the presentation having a piece of length L is at most $\ell^2(2m-1)^{2d\ell}(2m-1)^{-L}$ since there are $(2m-1)^{2d\ell}$ possible choices of couples of relators (we also have to check the special case when a relator shares a piece with itself, but this is not difficult). So if $L = \alpha\ell$ this makes $\ell^2(2m-1)^{(2d-\alpha)\ell}$. If $d < \alpha/2$, this tends to 0 as $\ell \to \infty$ (but all the more slowly as d is close to $\alpha/2$). One can reverse the argument to see that if $d > \alpha/2$, such an event actually occurs.

It is worth to compare this with Theorem 6. Indeed, when two relators share a piece of length $\alpha\ell$ we can form a van Kampen diagram D of boundary length $|\partial D| = 2\ell - 2\alpha\ell = |D|\,\ell(1-\alpha)$ so that this diagram contradicts Theorem 6 when $d < \alpha/2$.

The $C(p)$ condition amounts to the exclusion of a van Kampen diagram D in which a 2-cell is surrounded by $p-1$ 2-cells as on the left of Figure 1. Such a diagram D satisfies $|\partial D| = p\ell - 2\ell$ whereas Theorem 6 yields $|\partial D| \geqslant p\ell(1-2d-\varepsilon)$ so that (choosing $\varepsilon = (1/p-d)/10$) this is a contradiction when $d < 1/p$. This proves statement (3).

The $B(6)$ condition amounts to the exclusion of a diagram in which half the boundary of a 2-cell is covered by three other 2-cells as on the right of Figure 1. Note that this diagram D satisfies $|\partial D| = 4\ell - 2(\ell/2) = 3\ell$. Theorem 6 implies that $|\partial D| \geqslant 4\ell(1-2d-\varepsilon)$, so $d \geqslant 1/8 - \varepsilon/2$. So if $d < 1/8$ we get a contradiction (choosing e.g. $\varepsilon = (1/8-d)/10$). □

Remark 9.
By [Wis04], hyperbolicity and the $B(6)$ condition together imply the existence of a free and cocompact action on a CAT(0) cube complex. So this conclusion holds at density $< \frac{1}{8}$. This is a bit stronger than the $< \frac{1}{12}$ condition mentioned in the introduction.

Since the $C(6)$ condition is satisfied at density $< \frac{1}{6}$, our results suggest that generic $C(6)$ groups are a-T-menable. It is currently an open problem whether or not every infinite $C(6)$ group fails to satisfy Property (T).

In this paper we shall sometimes need to avoid some annoying topological configuration. This is the object of the next two propositions.

Proposition 10.
Let G be a random group at density $d < 1/4$. Let p be a closed path embedded in the Cayley graph of G. Then the length of p is at least ℓ; moreover, either p is the boundary path of some relator in the presentation, or the length of p is at least $\ell + \ell(1 - 4d - \varepsilon)$.

Consequently, the boundary paths of relators embed.

Proof.
This results from Theorem 6. Indeed, since p is not homotopic to 0, it is the boundary path of some van Kampen diagram D with at least one 2-cell, and so $|p| \geqslant \ell\,|D|\,(1 - 2d - \varepsilon)$. Now either $|D| = 1$ and p is the boundary path of a relator, or $|D| \geqslant 2$ and $|p| \geqslant 2\ell(1 - 2d - \varepsilon)$. □

Corollary 11.
Let G be a random group at density $d < 1/4$ and let $\widetilde{X}$ be the Cayley complex associated to the presentation. Let c_1, c_2 be two 2-cells in $\widetilde{X}$. Then $\partial c_1 \cap \partial c_2$ is connected.

Proof.
Suppose not and let v, w be two 0-cells of $\widetilde{X}$ lying in different components of $\partial c_1 \cap \partial c_2$. Let p_1, p_1' be the two paths in ∂c_1 joining v to w on each side of c_1, and likewise let $p_2, p'2$ be the two paths in ∂c_2 joining v to w.

Each of the paths $p_1 p_2^{-1}, p_1 {p_2'}^{-1}, p_1' p_2^{-1}$ and $p_1' {p_2'}^{-1}$ is a closed path in the 1-skeleton of $\tilde{X}$. Each of these paths is not null-homotopic in this 1-skeleton, otherwise v and w would lie in the same component of $\partial c_1 \cap \partial c_2$. So by Proposition 10 each of these paths has length at least ℓ, and since $|p_1| + |p_1'| = |p_2| + |p_2'| = \ell$, the only possibility is that $|p_1| = |p_1'| = |p_2| = |p_2'| = \ell/2$. This implies that $\left|p_1 p_2^{-1}\right| = \ell$, so that $p_1 p_2^{-1}$ is the boundary path of some 2-cell c_3. Now c_1 and c_3 share half of their boundary length, which at $d < 1/4$ contradicts Proposition 8. □

Another notion we shall need is that of *fulfilling* of a diagram. Let D, an *abstract diagram*, be a finite connected graph embedded in the plane, each edge of which is decorated with a positive integer, its *length*. Let $\langle\, a_1, \ldots, a_m \mid R\,\rangle$ be any group presentation. A *fulfilling* of D is the attribution to each face of D of a relator in R (together with an orientation) such that the resulting object is a reduced van Kampen diagram of the presentation, in a way compatible with the prescribed lengths (see the notion of *decorated abstract van Kampen diagram* in [Oll04] for precisions).

The following appears in [Oll05a], Propositions 12 and 13:

Theorem 12.
Let $G = \langle\, a_1, \ldots, a_m \mid R\,\rangle$ be a group presentation. For any abstract diagram D, let $S_n(D)$ be the number of n-tuples of distinct relators in R such that there exists a

fulfilling of D using these relators (n is at most the number of faces $|D|$ of D since a relator may be used multiple times in the diagram).

For random groups at density d, for any abstract diagram D we have the following bound on the expectation of $S_n(D)$:

$$\mathbb{E}S_n(D) \leqslant (2m-1)^{\frac{1}{2}(|\partial D|-(1-2d)\ell|D|)}$$

and so for any D, with overwhelming probability we have:

$$S_n(D) \leqslant (2m-1)^{\frac{1}{2}(|\partial D|-(1-2d-\varepsilon)\ell|D|)}$$

We note that the second assertion in Theorem 12 (which holds for fixed D) follows from the first one by the Markov inequality.

2 Hypergraphs and carriers

2.1 Historical background on cubulating groups

The results in this paper employ Sageev's construction [Sag95] of an action on a CAT(0) cube complex from a group G and a codimension-1 subgroup H.

Niblo and Reeves [NR03] and Wise [Wis04] had observed that Sageev's construction works in the context of "geometric spaces with walls". For Coxeter groups, these walls are the reflection walls stabilized by the involutions in the Coxeter complex. For small cancellation groups, the walls are constructed as we do here: by producing immersed graphs in a 2-complex that are transverse to the 1-skeleton and such that each edge of the graph bisects a 2-cell. The walls corresponding to such graphs appear to have played a role in Ballmann-Swiątkowski's proof of the failure of Property (T) for the geometric case of $(4,4)$-complexes and $(6,3)$-complexes [BŚ97].

It is clear from [Wis04] that Sageev's cubulation result can be carried out for a family of more general codimension-1 graphs which embed, are transverse to the 1-skeleton, and locally separate the 2-complex. These are examples of Dunwoody's "tracks" and we expect they will be referred to as "walls" in future work on this subject. Indeed, subsequently, Nica [Nic04] and Chatterji and Niblo [CN04] have written out an explicit application of Sageev's construction to cubulate abstract "spaces with walls". Those were introduced by Haglund and Paulin [HP98] especially motivated by Coxeter groups and CAT(0) cube complexes.

Building upon [Sag97, NR03, Wis04], Hruska and Wise [HW04] have laid out "axioms" on a space with walls (or 2-complex with hypergraphs) for verifying finiteness properties of the cubulation. We follow the framework there to verify our main results. We expect there will be further work along these lines.

2.2 Definition of hypergraphs

In a nutshell, hypergraphs in a 2-complex are obtained by drawing a segment between the midpoints of each pair of opposite 1-cells in each 2-cell. These segments define a graph, the connected components of which are the hypergraphs. We give a more precise definition below.

Definition 13.
Let $\widetilde{X}$ be a simply connected 2-complex. We suppose that each 2-cell of X has even boundary length. (If this is not the case, we just perform a subdivision of all 1-edges of $\widetilde{X}$ before constructing hypergraphs.)

We define a graph Γ as follows: The set of vertices of Γ is the set of 1-cells of $\widetilde{X}$. There is an edge in Γ between two vertices if there is some 2-cell R of X such that these vertices correspond to antipodal 1-cells in the boundary of R (if there are several such 2-cells R, we put as many edges in Γ). The 2-cell R is the 2-cell of $\widetilde{X}$ containing *the edge.*

There is a natural map φ from Γ to a geometric realization of it in $\widetilde{X}$, which sends each vertex of Γ to the midpoint of the corresponding 1-cell of X, and each edge of Γ to a segment joining two antipodal points in the 2-cell R. Note that the images of two edges contained in the same 2-cell R always intersect, so that in general φ is not an embedding.

Let $\Lambda_i, i \in I$ be the connected components of Γ, and let $\widetilde{\Lambda}_i, i \in I$ be their universal covers, which are trees. A hypergraph *in $\widetilde{X}$ is any of the maps $\widetilde{\varphi} : \widetilde{\Lambda}_i \to \widetilde{X}$. We will often write "the hypergraph Λ_i" to denote this map. The 1-cells of $\widetilde{X}$ through which a hypergraph passes are* dual *to it. The hypergraph* embeds *if $\widetilde{\varphi}$ is an embedding, i.e. if the image of the hypergraph in $\widetilde{X}$ is a non-self-crossing tree.*

For each subgraph $A \subset \widetilde{\Gamma}$, we define a 2-complex V, the unfolded carrier *of A, in the following way: For each edge e in A contained in the 2-cell R of $\widetilde{X}$, consider an isomorphic copy R_e of R. Now take the disjoint union of these copies and glue them as follows: if edges e and e' of A share a common endpoint $v \in A$, identify R_e and $R_{e'}$ along the 1-cell corresponding to vertex v. When A is connected, it is by construction an embedded hypergraph of its unfolded carrier.*

A hypergraph segment *(resp.* ray*, resp.* line*) is an immersed finite path (resp. immersed ray, immersed line) in a hypergraph. A* ladder *is the unfolded carrier of a segment.*

Remark 14.
The term "hypergraph" is a misnomer, which arose as a graph corresponding to a "hyperplane" in a CAT(0) cube complex C. Hypergraphs will play the role of "codimension-1 subgraphs" later in the paper. The term hypergraph is used in graph theory to mean a certain high-dimensional generalization of a graph, but we will have no use for that notion in this paper.

Lemma 15.
Suppose a hypergraph Λ embeds in the simply connected complex $\widetilde{X}$. Then $\widetilde{X} - \Lambda$ consists of two components.

Proof.
This follows easily from the fact that $\mathsf{H}_1(\widetilde{X}) = 0$ and a Mayer-Vietoris sequence

argument applied to the complement of the hypergraph and a neighborhood of the hypergraph. □

3 Studying hypergraphs with collared diagrams

In this section we define and examine various notions of "collared diagrams". In Section 3.1, we show that hypergraphs are trees unless certain collared diagrams exist. In Section 3.3, we show that the intersection of a pair of hypergraphs contains at most one point, unless there is a certain collared diagram between them. In Section 3.4, we explain that if a geodesic touches a hypergraph in exactly two points, then there is a certain relatively collared diagram between the geodesic and the hypergraph. In each case, various quasicollared diagrams will serve as a useful technical object to facilitate the proofs.

Convention 16 (Conventions on $\widetilde{X}$ and its hypergraphs).
In the remainder of the paper, $\widetilde{X}$ is the Cayley 2-complex of a random group (and hence all relations have the same length). However in this section we work under more general hypotheses (which the reader is welcome to ignore). Our only hypothesis on $\widetilde{X}$ is that it is a simply connected combinatorial 2-complex, and that the boundary cycle of each 2-cell is an immersed path in $\widetilde{X}^1$, of even length.

3.1 Collared diagrams

We refer to [MW02] (Def. 2.6) for the definition of *disc diagrams*, which play for arbitrary 2-complexes the role of van Kampen diagrams for Cayley complexes. The reader may just read "van Kampen diagram".

The central notion in this section is the following (see Figure 2):

Definition 17 (Collared diagram).
A collared diagram *is a disc diagram $D \to \widetilde{X}$ with the following properties:*

1. *there is an external 2-cell C called a* corner *of D*
2. *there is a hypergraph segment $\lambda \to D \to \widetilde{X}$ of length at least 2*
3. *the first and last edge of λ lie in C, and no other edge lies in C*
4. *λ passes through every other external 2-cell of D exactly once*
5. *λ does not pass through any internal 2-cell of D.*

D is cornerless *if moreover the first and last edge of λ coincide in C (in which case the hypergraph cycles).*

The above definition implies that the diagram is homeomorphic to a disc.

Remark 18.
Note that we do not exclude that C is the only boundary edge of D (in which case the

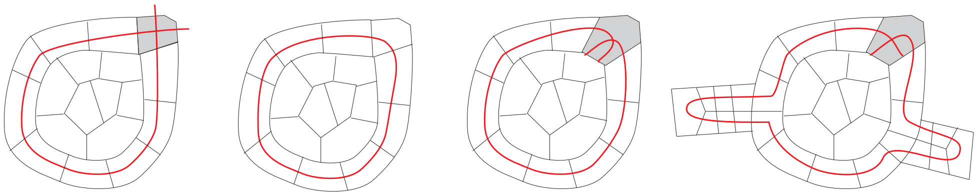


Figure 2: Several kinds of collared diagrams: The corner 2-cell of the first collared diagram on is shaded. The second collared diagram is cornerless. The hypergraph segment of the third collared diagram ends before it enters the interior. The last collared diagram is more typical in the sense that the carrier of the segment folds.

boundary path of C is not simple). However, since the boundary path of any 2-cell is immersed by assumption, D has at least two 2-cells. But it might not have any internal 2-cells.

Definition 19.
A cancellable pair *in $Y \to X$ is a pair of distinct 2-cells R_1, R_2 meeting along an edge e in Y such that R_1 and R_2 map to the same 2-cell in X, and moreover, the boundary paths of R_1 and R_2 starting at e, map to the same path in X. A map $Y \to X$ is* reduced *if Y contains no cancellable pairs. Note that the composition of reduced maps is reduced.*

In our framework, $Y \to X$ is reduced precisely if $Y \to X$ is a near-immersion *meaning that $(Y - Y^0) \to X$ is an immersion. See [MW02] for more about reduced maps.*

For van Kampen diagrams this notion coincides with the usual notion of reduced diagram (at least if relators which are proper powers are handled correctly, which is a messy point in the van Kampen diagram literature).

The main goal of this section is to prove the following theorem.

Theorem 20.
Let Λ be some hypergraph. The following are equivalent:

1. *Λ embeds.*
2. *There is no reduced collared diagram collared by a segment of Λ.*
3. *There is no quasicollared diagram collared by a segment of Λ (Definition 21).*

Proof.
If there is a reduced diagram $E \to \widetilde{X}$ collared by a segment λ of Λ then clearly $\Lambda \to \widetilde{X}$ is not an embedding. Indeed, the path $\lambda \to E$ has the property that its first and last edges cross or coincide, and so this is the case for $\lambda \to \widetilde{X}$.

The converse, which plays an important role in this paper needs a bit more work, and so we outline the proof which employs several lemmas proven later in this section. Suppose Λ does not embed.

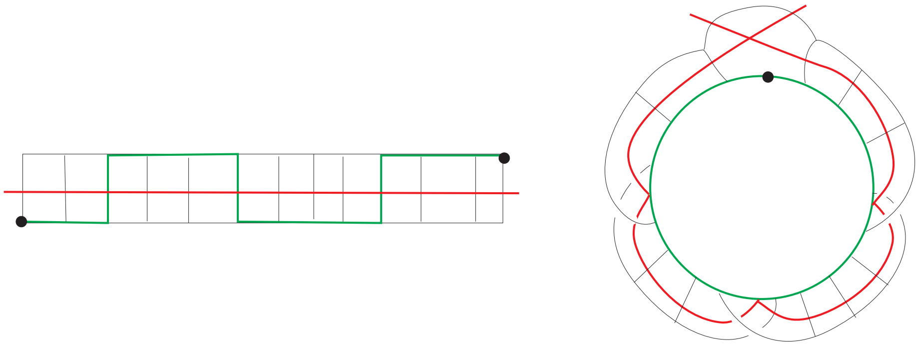


Figure 3: On the left is the path P in the ladder L containing part of the hypergraph Λ. On the right is the quasicollared disc diagram $L \cup_P D$ obtained by attaching L to D along P.

In Lemma 23, we prove that if Λ doesn't embed in $\widetilde{X}$, then there exists a diagram quasicollared by Λ (Definition 21), denoted by $F \to \widetilde{X}$.

In Lemma 24, we show that by removing cancellable pairs, we can assume that $F \to \widetilde{X}$ is reduced.

In Lemma 25 we extract a reduced collared diagram $E \to \widetilde{X}$ from the reduced quasicollared diagram $F \to \widetilde{X}$. □

We now define quasicollared diagrams which, unlike collared diagrams, do not have an easily stated intrinsic definition.

Definition 21 (Quasicollared diagram).
Consider the ladder L of some hypergraph segment λ of length at least 2. We suppose that the first and last 2-cells C_1, C_2 of L map to the same two-cell of $\widetilde{X}$.

Let $A = L/\{C_1{=}C_2\}$ be the complex obtained from L by identifying the closures of C_1 and C_2.

Let $P \to A$ be a simple cycle in P representing a generator of $\mathsf{H}_1(A)$. Suppose that there exists a disc diagram $D \to \widetilde{X}$ with boundary path P.

A quasicollared diagram $F \to X$ is the complex obtained by forming the union $F = A \cup_P D$. (See Figure 3.)

Remark 22.
F is a genuine disc diagram precisely when $P \to A$ does not cross λ. This happens precisely when A is a cylinder instead of a Moebius strip and $P \to A$ is a boundary cycle of A.

Lemma 23 (Existence).
Suppose that the hypergraph Λ does not embed in $\widetilde{X}$. Then there exists a quasicollared diagram $F \to \widetilde{X}$ that is collared by Λ.

Proof.
Suppose the hypergraph Λ contains a nontrivial immersed edge-path $\lambda \to \Lambda$ such that

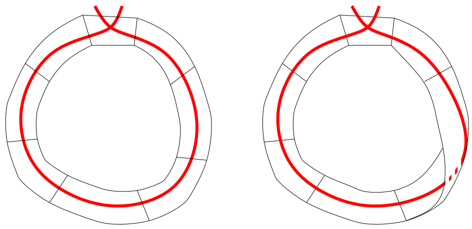


Figure 4: The basic loops.

λ projects to a non-simple path in $\widetilde{X}$. We can assume that λ is minimal, that is, any proper subsegment of λ embeds.

So the 2-cells of $\widetilde{X}$ containing the first and last edge of λ are the same. Let L be the ladder carrying λ; its first and last 2-cells C_1, C_2 map to the same 2-cell of $\widetilde{X}$.

We can therefore form a quotient space $A = L/\{C_1{=}C_2\}$ and there is an induced map $A \to \widetilde{X}$. We will refer to the cell $C_1 = C_2$ as the *corner*. As in Figure 4, there are two cases for A according to whether or not λ "preserves orientation" of the ladder.

Let $P \to A$ be a simple cycle in A that maps to a generator of $\pi_1(A)$. Since $\widetilde{X}$ is simply connected, there is a disc diagram $D \to \widetilde{X}$ whose boundary path is P.

Note that while $P \to A$ is an immersion, the map $A \to \widetilde{X}$ may not be, and so it is possible that D is singular, and may have spurs.

Finally we form the desired quasicollared diagram $F = A \cup_P D$.

If we think of P as a path in L instead of A, then P may travel from one side of L to the other, as in Figure 3. Indeed, this is always the case in the orientation reversing case where A is a Moebius strip. F is a genuine disc diagram exactly when λ does not cross any edge of P. $\square$

Lemma 24 (Reducing).
Let $F \to X$ be a quasicollared diagram. Then there exists a reduced quasicollared diagram $F' \to X$ which is collared by a subsegment of the hypergraph segment collaring F.

Proof.
Keeping the notation in the definition of quasicollared diagrams, there are three types of cancellable pairs in $F \to \widetilde{X}$ to consider according to whether the 2-cells lie in: D, D or D, A, or A, A.

In the first case, the cancellable pair is removed in the usual way for van Kampen diagrams (prone to errors in the literature, but works nevertheless...): We remove the open 2-cells and the open 1-cell along which they form a cancellable pair, and we identify their remaining corresponding boundaries.

In the second case, we can adjust our choice of P to form a new simple cycle. Namely let C_1, C_2 be the 2-cells forming the cancellable pair, with $C_1 \subset A$ and $C_2 \subset D$. Push P across to the other side of C_1. Now C_2 can be removed from D. This is illustrated as the first two configurations in Figure 5, where cancellable pairs are marked by dots: the first configuration is the case when originally P does not jump

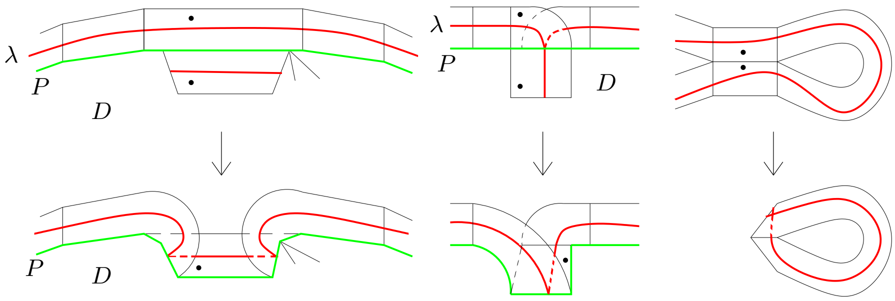


Figure 5: A cancellable pair between D and A, another cancellable pair in a slightly different position, and a cancellable pair between A and A. The cancellable pairs are marked with dots.

from one side of A to the other around some side of C_1 (in which case the new P crosses A twice at this point); the second configuration is when originally P crosses A along some side of C_1, in which case the new P crosses A along the other side of C_1 afterwards.

In the third case (reduction between A and A), this means that there is a pair of 2-cells C_1, C_2 in L, different from the pair of extremal 2-cells, mapping to the same 2-cell of $\widetilde{X}$. This means that we can find a proper subsegment λ' of the hypergraph segment λ which does not embed in $\widetilde{X}$. The ladder carrying λ', which has C_1 and C_2 as extremal 2-cells, can now be used to define a smaller quasicollared diagram as in the rightmost illustration of Figure 5.

Keep reducing cancellable pairs. The only thing to check is that eventually A is not empty. Observe that reductions between D and D and between D and A preserve A and λ. So the only way A can become empty is if at some step two *consecutive* 2-cells of A are cancellable. But this means that λ was not immersed, which it is by definition of a hypergraph segment. □

We now extract a collared diagram from a quasicollared one.

Lemma 25 (Collaring).
If there is a reduced quasicollared diagram F, then there is a reduced collared diagram F'. Moreover, F' and F are collared by segments of the same hypergraph of $\widetilde{X}$.

Proof.
Keeping the same notation again, suppose some edge e of P crosses Λ, that is, consider an edge e in P that is dual to Λ (witnessing for the fact that the diagram is quasicollared but not collared). Observe that Λ enters D at e. Let λ' be the path of Λ in D issuing from e. As on the left in Figure 6, either λ' is simple or λ' crosses itself in D.

If λ' crosses itself then we choose some subpath λ'' of λ' that is a simple loop in D bounding some topological disc in D, as in the middle diagram of Figure 6.

There is then a diagram D' having λ'' as the hypergraph in its collar. This is illustrated on the right in Figure 6. (Note that D' is only nearly a subdiagram of D

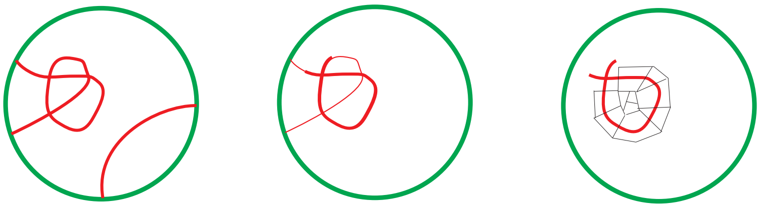


Figure 6:

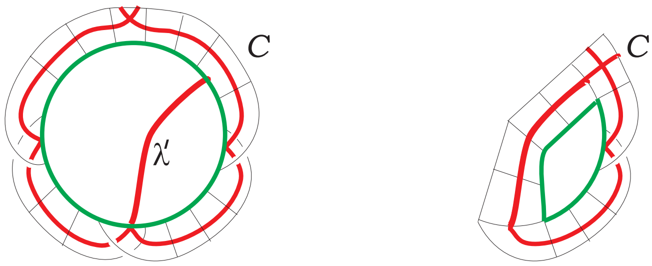


Figure 7:

since the map $D' \to D$ might fail to be injective on $\partial D'$.)

The other possibility is that the path λ' is simple in D (Figure 7). In this case, λ' has to exit D by crossing the collar at some 2-cell C, dividing F into two halves. Pick the half of F that does not contain the corner of F: this provides a new quasicollared diagram F' with C as its corner. (If C happens to be the corner of F already, then any half will do.)

This new diagram is smaller than F in the sense that the number of intersections between ∂D and the hypergraph segment collaring the diagram decreases. So repeating the process will eventually provide a collared diagram. The last step is illustrated on Figure 8.

The new diagram obtained is reduced since F itself is. □

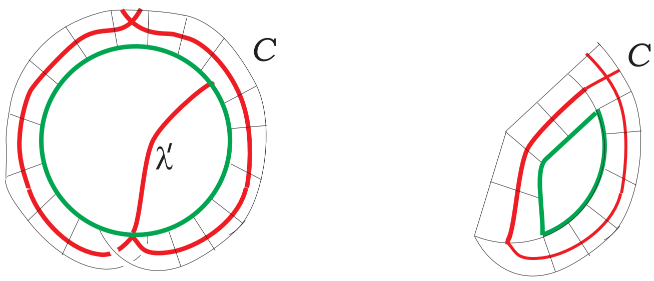


Figure 8:

3.2 Diagrams quasicollared by hypergraphs and paths.

We now give a definition of a notion generalizing that of quasicollared diagram, in which we allow the collar to consist of segments of several hypergraphs and/or paths in $\widetilde{X}$.

Definition 26.
Let $n \geqslant 2$ be an integer and decompose $\{1, \ldots, n\}$ as a disjoint union $I \cup J$ (where I or J may be empty). For $i \in I$ let λ_i be a hypergraph segment of length at least 2, carried in ladder L_i. For $i \in I$ let also p_i be a path immersed in L_i joining a point in the boundary of the first 2-cell of L_i and a point in the boundary of the last 2-cell of L_i. For $j \in J$ let p_j be any path immersed in $\widetilde{X}$.

We suppose that (subscripts mod n):

1. *When $i \in I$ and $i+1 \in I$, then: The final 2-cell of L_i and the initial 2-cell of L_{i+1} have the same image in $\widetilde{X}$. Moreover the final edge of λ_i and the initial edge of λ_{i+1} do not have the same image in $\widetilde{X}$. Moreover the (images in $\widetilde{X}$ of) initial point of p_{i+1} and the final point of p_i coincide.*

2. *When $i \in I$ and $i+1 \in J$, then the image in $\widetilde{X}$ of the final point of p_i coincides with the initial point of p_{i+1}, and likewise when $i \in J$ and $i+1 \in I$.*

3. *If $i \in J$ then both $i+1$ and $i-1$ lie in I.*

This allows to define a cyclic path $P = \cup p_i$. Let D be a disc diagram with boundary path P.

Let A' be the disjoint union of L_i for $i \in I$ and let A be the quotient of A' under the identification of the last 2-cell of L_i with the first 2-cell of L_{i+1} whenever $i, i+1 \in I$.

A diagram quasicollared *by the $\lambda_i, i \in I$ and the $p_j, j \in J$ is the union $E = D \cup_{p_i, i \in I} A$.*

The corners *of E are the initial and final 2-cells of the L_i's.*

It is said to be collared *by the $\lambda_i, i \in I$* and *the $p_j, j \in J$ if moreover, for $i \in I$, the path p_i does not cross λ_i except maybe at the endpoints of λ_i. In this case E is a genuine disc diagram.*

We say that the hypergraphs λ_i do not enter *E if for $i \in I$, the initial and final points of λ_i belong to the boundary of E.*

Note that D may be singular, and may even contain no 2-cell in the case P is a null-homotopic path in the 1-skeleton of $\widetilde{X}$. Note also that we do not allow "cornerless" diagrams since we imposed that two successive hypergraph segments intersect transversely (otherwise, simply unite them).

3.3 2-collared diagrams

A 2-*collared diagram* is a diagram collared by two hypergraph segments.

The main goal of this section is the following:

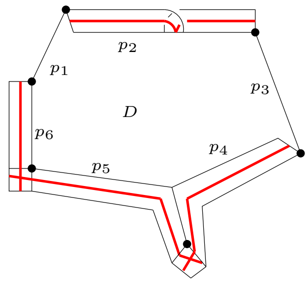


Figure 9: A diagram collared by hypergraphs and paths.

Theorem 27.
Suppose Λ_1 and Λ_2 are distinct hypergraphs that embed in $\widetilde{X}$. There is more than one point in $\Lambda_1 \cap \Lambda_2$ if and only if there exists a reduced diagram E collared by segments of Λ_1 and Λ_2. Moreover, if Λ_1 and Λ_2 cross at a 2-cell C, then we can choose E so that C is one of its corners.

Proof.
If there exists a diagram $E \to \widetilde{X}$ collared by Λ_1 and Λ_2, then the intersections of λ_1 and λ_2 in the two corners 2-cells map to two intersection points in $\widetilde{X}$, which are distinct since Λ_1 and Λ_2 are embedded.

The converse requires more work, and we outline its proof which depends upon lemmas proven in this section.

In Lemma 28, we show that if Λ_1 and Λ_2 intersect twice, then there is a quasicollared diagram between them.

In Lemma 30, we show that if there is a quasicollared diagram between them then there is a reduced quasicollared diagram between them.

In Lemma 31, we extract a reduced collared diagram between Λ_1 and Λ_2, from a reduced quasicollared diagram. □

Lemma 28 (Existence of 2-quasicollared diagrams).
Suppose there are distinct embedded hypergraphs Λ_1 and Λ_2 whose images in $\widetilde{X}$ intersect in more than one point. Then there exists a diagram F quasicollared by Λ_1 and Λ_2.

Proof.
Let λ_1, λ_2 be hypergraph segments in Λ_1, Λ_2 which intersect at the centers of their first and last edges. Let L_i be the ladder carrying λ_i, and observe that the first and last 2-cells of L_1, L_2 project to the same 2-cells of $\widetilde{X}$. Let $A \to \widetilde{X}$ be obtained by forming the union of L_1 and L_2 and identifying their first and last closed 2-cells. Observe that $\pi_1(A) \cong \mathbb{Z}$ except for the degenerate case where each L_i consists entirely of these first and last 2-cells.

In this degenerate case, define $F = A$. Otherwise, let $P \to A$ be a simple closed path representing a generator of $\pi_1(A)$. Let $D \to \widetilde{X}$ be a disc diagram with boundary path $P \to \widetilde{X}$. Let $F = A \cup_P D$. □

Lemma 29.
Let F be a diagram quasicollared by two embedded hypergraph segments λ_1, λ_2. Then there exists a diagram F' quasicollared by two subsegments λ'_1, λ'_2 of λ_1, λ_2, such that the only 2-cells in the intersection of the images of the ladders of λ'_1 and λ'_2 in $\widetilde{X}$ are the corners of F'. Moreover, F' can be chosen to contain either corner of F as one of its corners.

Proof.
This is more difficult to state than to prove. Let C_1 be the first corner of F. Let C_2 be the first 2-cell in the ladder of λ_1, distinct from C_1, which lies in the image of the ladder of λ_2 in $\widetilde{X}$. Taking the corresponding initial subsegments of λ_1 and λ_2 and applying Lemma 28, we get what we need, preserving corner C_1. □

Lemma 30 (Reducing).
Let F be a diagram quasicollared by two embedded hypergraphs. Then there exists a reduced diagram F' quasicollared by two segments of the same hypergraphs, and moreover F' can be chosen to contain either corner of F.

Proof.
This is similar to Lemma 24. Applying Lemma 29, we can suppose that the images of the ladders collaring F intersect only at the two corners of F.

Keep the notation of Definition 26. Cancellable pairs between 2-cells in D, D can be removed as usual to lower the total number of 2-cells. If there is a cancellable pair between D and A then by pushing the boundary path across the 2-cell in A, we obtain a new simple cycle with a smaller disc diagram as in Lemma 24.

Now for the cases of a cancellable pair between A and A. The cancellable pair cannot lie between 2-cells in the same ladder L_i, for this would imply that λ_i does not embed. It cannot lie between L_1 and L_2 either since this would contradict the conclusion of Lemma 29. □

Lemma 31 (Collaring).
Let F be a reduced diagram quasicollared by two embedded hypergraphs Λ_1, Λ_2. Then there is a reduced diagram F' collared by Λ_1 and Λ_2. Moreover, F' can be chosen to share either corner with F.

Proof.
Keeping notation again, suppose that F is quasicollared but not collared, i.e. that P crosses the segment λ_1. (The argument for λ_2 is identical.) Consider the first such situation on λ_1. Then λ_1 can be extended into a hypergraph segment μ_1 that enters D. μ_1 cannot cross itself since Λ_1 embeds. So μ_1 exits D by crossing λ_2. Then by choosing the part of F lying between μ_1 and λ_2 we get a diagram which is quasicollared by λ_2 and μ_1, containing the first corner of F. Repeating the argument with λ_2 ends the proof. □

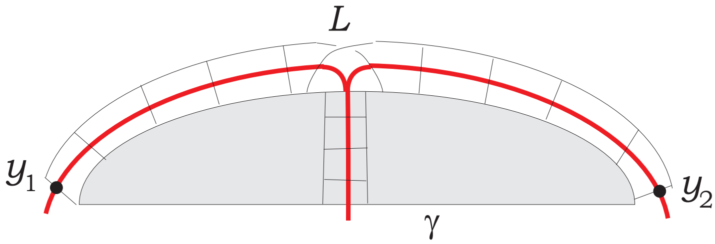


Figure 10:

3.4 Diagrams collared by a hypergraph and a path

Lemma 32 (Existence).
Let Λ be a hypergraph which embeds in $\widetilde{X}$. Let λ be a segment of Λ. Let γ be an embedded path in $\widetilde{X}$ with the same endpoints as λ. (Here γ is an edge path which starts and ends at "mid-edge vertices" corresponding to vertices of Λ.) Then there exists a reduced diagram F quasicollared by λ and γ.

Moreover, if γ does not intersect Λ anywhere except at its endpoints, then F is actually collared.

Remark 33.
γ will be geodesic in our applications. This will serve to study the metric properties of embedded hypergraphs.

Proof.
Proceed exactly as above to get a reduced diagram quasicollared by λ and γ: Let L be the ladder carrying λ. Let $P \to L$ be a simple edge-path with the same endpoints as λ. Let $D \to \widetilde{X}$ be a disc diagram with boundary path $P^{-1}\gamma$. The 2-complex $F = D \cup_P L$ is the desired quasicollared diagram between Λ and γ.

The reduction process is carried out as above. Note that since λ embeds there is no pair of cancellable 2-cells between L and L, and so λ is preserved by the reduction process.

Now suppose that F is not a *collared* diagram. Then λ passes through an edge of P. Thus λ can be extended into D by a segment μ in D lying in the same hypergraph Λ (see Figure 10). Since Λ is embedded, μ cannot exit D by crossing λ. So it has to exit D by crossing γ, thus providing a third intersection point between Λ and γ, contrary to the assumption. $\square$

4 The hypergraphs are embedded trees at $d < 1/5$

Henceforth, $\widetilde{X}$ is the Cayley 2-complex associated to a finite presentation of the random group G at density d and length ℓ (Def. 1), that is, $\widetilde{X}$ is the universal cover of the standard 2-complex associated to the presentation.

Definition 34.
Let D be a van Kampen diagram. The external 1-cells *of D are the 1-cells which lie in ∂D. The other 1-cells are* internal. *A 2-cell of D is* external *if its closure contains an external 1-cell, otherwise it is* internal.

A shell *of D is a 2-cell R such that the boundary path of D contains a subpath Q, where Q is a subpath of the boundary path of R, and $|Q| > \frac{1}{2}|\partial R|$.*

A spur *of D is an 1-cell ending at a valence 1 0-cell on ∂D. Note that D has no spur if and only if its boundary path is immersed.*

The following frequently arising condition is a special case of Greendlinger's lemma for $C'(\frac{1}{6})$ presentations:

Condition 35.
For every reduced spurless van Kampen diagram $D \to X$ either

1. *D has at most one 2-cell.*
2. *D contains at least two shells.*

Theorem 36.
Let X be the standard 2-complex of some presentation. Suppose that X satisfies Condition 35. Then there is no reduced collared van Kampen diagram $D \to X$ (either cornerless or with a corner)

Consequently, all hypergraphs are trees embedded in $\widetilde{X}$.

Proof.
We show that there is no collared diagram. Indeed, suppose there is a collared diagram. It has no spurs, and has more than one 2-cell. But its only possible shell is its corner. Indeed, every other external 2-shell R, contains an edge of a hypergraph in the interior of D, so any path on $\partial R \cap \partial D$ has length $< \frac{1}{2}|\partial R|$. This contradicts Condition 35. □

Corollary 37.
For random groups at density $d < 1/5$, with overwhelming probability all hypergraphs embed in $\widetilde{X}$.

Proof.
Theorem 6 in [Oll07] states that Condition 35 holds with overwhelming probability for random groups at density $d < 1/5$. We can therefore apply Theorem 36. □

The goal of section 11 is to prove that as soon as $d > 1/5$, on the contrary there is only one hypergraph, which crosses every 1-cell of $\widetilde{X}$. Figure 21 at the end of the paper shows why hypergraphs do not embed at $d > 1/5$.

We now turn to the metric aspect of the embeddings.

Theorem 38.
Consider a random group at density $d < 1/5$. With overwhelming probability, the distance in $\widetilde{X}^1$ between two vertices of a hypergraph Λ is at least $(1/2 - 2d - \varepsilon)\ell$ times the minimal number of edges joining them in Λ.

Proof.
Let γ be a geodesic in $\widetilde{X}^1$ between two points $y_1, y_2 \in \Lambda$. It is sufficient to prove the statement of the theorem under the additional hypothesis that γ does not intersect Λ at any other points.

By Lemma 32, there exists a reduced diagram E collared by γ and a ladder L carrying a segment of Λ.

Since E is collared and not only quasicollared, in particular it is an ordinary van Kampen diagram. Let n be the number of cells in the ladder L. We have $|\partial E| \leqslant n\ell/2 + |\gamma|$. But by Theorem 6, up to some ε we have $|\partial E| \geqslant (1-2d)\ell\,|E| \geqslant (1-2d)n\ell$ and so as claimed we have:

$$|\gamma| \geqslant n\ell(\frac{1}{2} - 2d)$$

□

Note that the multiplicative constant does not vanish as $d \to \frac{1}{5}$ (compare Figure 17 at $d > \frac{1}{5}$). However in the proof of this theorem, we already used that hypergraphs embed (a condition needed in our previous study of diagrams collared between a hypergraph and a geodesic).

Corollary 39.
In random groups at density $d < \frac{1}{5}$, with overwhelming probability, the stabilizer of any hypergraph is a free, quasiconvex subgroup.

Proof.
Since hypergraphs are trees in the Cayley complex, their stabilizers acts freely on a tree. Since random groups are torsion-free, so are the stabilizers, hence freeness since torsion-free groups acting freely on trees are free. Now a quasi-isometrically embedded tree in a hyperbolic space is quasiconvex, since quasi-geodesics remain at bounded distance from geodesics. □

5 2-collared diagrams

In Section 4, we showed that hypergraphs do not self-intersect at density $d < \frac{1}{5}$. It will also be useful to understand the way a pair of hypergraphs can intersect each other. A naive hope would be that distinct hypergraphs are either disjoint or intersect in a single point, but this is almost never the case as Figure 11 shows. However, we will show that at low density, intersecting hypergraphs might “braid” with each other a bit, but after departing do not converge again, so that intersection is a relatively local matter.

In Theorem 36 we saw that there are no 1-quasicollared diagrams at $d < \frac{1}{5}$. We now turn to 2-collared diagrams.

Theorem 40.
For random groups at $d < \frac{1}{6}$, with overwhelming probability every reduced diagram with at least three 2-cells has at least three shells.

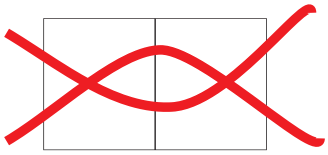


Figure 11:

In particular, there exists no reduced 2-collared diagram except the one depicted in Figure 11.

Corollary 41.
For random groups at density $d < \frac{1}{6}$, with overwhelming probability the following holds: Let Λ_1, Λ_2 be two hypergraph rays intersecting in 2-cell c. Either they intersect in a 2-cell adjacent to c as in Figure 11, or they do not intersect anywhere else.

Proof of the corollary.
This follows from the theorem by Theorem 27. □

To prove the theorem, we shall need the following lemma (which will be of independent use).

Lemma 42.
Consider a random group at density $d < 1/4$. With overwhelming probability the following holds.

Let D be a reduced spurless van Kampen diagram. Let k be the number of shells in D. Let p be the number of internal 2-cells of D.

Then the number of external 2-cells in D is at most

$$\frac{(1/2-d)(k/2-p)}{1/4-d}+\varepsilon$$

Proof of the lemma.
Let A be the set of 2-cells R of D with $\partial R \cap \partial D \neq \varnothing$. Let $B \subset A$ be those 2-cells R with $|\partial R \cap \partial D| \geqslant \ell/2$ (the shells of D). Let $C \subset B$ be those 2-cells R with $|\partial R \cap \partial D| \geqslant \ell(1-d)$. Let $n = \#A$, $k = \#B$ and $q = \#C$.

Let D' be the diagram obtained from D by removing the 2-cells in C. (D' might not be connected, but this does not matter since Theorem 6 applies to non-connected diagrams as well.) Let us evaluate the boundary length of D'. By definition, a 2-cell in C contributes at most $d\ell$ edges to $\partial D'$ (the ones that were not on ∂D). All other edges of $\partial D'$ were already present on ∂D and belonged to the boundary of a 2-cell in $A - C$. A 2-cell in $A - B$ contributes at most $\ell/2$ edges and a 2-cell in $B - C$ contributes at most $\ell(1-d)$ edges. So we have

$$\left|\partial D'\right| \leqslant qd\ell + (n-k)\frac{\ell}{2} + (k-q)(1-d)\ell.$$

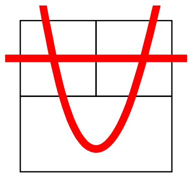


Figure 12:

On the other hand, by Theorem 6,

$$\left|\partial D'\right| \geqslant (1-2d-\varepsilon)\ell \left|D'\right| = (1-2d-\varepsilon)(n+p-q)\ell$$

and the combination of these two inequalities yields the conclusion. □

Proof of Theorem 40.
Suppose there are at most two shells. By Lemma 42, the number of external 2-cells is bounded above by $\frac{(1/2-d)(k/2-p)}{1/4-d}+\varepsilon$. When $k=2$ and $d<\frac{1}{6}$ this bound is <4 and so there are at most 3 external 2-cells. Note that if $p\geqslant 1$ then there are $<\varepsilon$ external 2-cells, hence $p=0$ and there are no internal 2-cells.

So it is enough to rule out diagrams D having exactly three 2-cells r_1, r_2, r_3 where only r_1, r_2 are shells. Since r_3 is not a shell the internal length of D is more than $\ell/2$ and so $|\partial D| < |D|\,\ell - 2(\ell/2) = 2\ell$. But for $d<1/6$ Theorem 6 yields $|\partial D| \geqslant \frac{2}{3}\,\ell\,|D| = 2\ell$ (choosing e.g. $\varepsilon < (1/6-d)/10$) hence a contradiction.

Note that a 2-collared diagram has at most two shells (its corners). □

Another consequence of the lemma is the following.

Theorem 43.
For random groups at $d<\frac{1}{5}$, with overwhelming probability, any reduced 2-collared diagram has at most five 2-cells, and no internal cells.

Proof.
For $d<\frac{1}{5}$ and $k=2$, and small enough ε (depending on $1/5-d$), the quantity $\frac{(1/2-d)(k/2-p)}{1/4-d}+\varepsilon$ is less than 6, and less than ε if $p\geqslant 1$. □

A less sharp version of this last assertion probably follows from the quasiconvexity obtained in Section 4.

6 Typical carrier 2-cells at $d<1/5$

We saw in Theorem 40, that at densities less than 1/6, there are no nondegenerate 2-collared diagrams, whereas it is not difficult to check that there are some at densities between 1/6 and 1/5 (e.g. the one of Figure 12). However, as demonstrated in this section, for "most" 2-cells of $\widetilde{X}$, there are no 2-collared diagrams having these 2-cells as corner cells.

Let $(r_1, \ldots, r_N)$ be the N-tuple of random relators making the presentation, where by definition $N = (2m-1)^{d\ell}$. In the sequel we prove that some bad properties are excluded for relator r_1 with overwhelming probability (these properties are excluded with high probability for any relator r_i with i *fixed in advance*; however, for any random sample $(r_1, \ldots, r_N)$, there might be some i depending on the random sample, such that r_i satisfies these bad properties).

Lemma 44.
Consider a random group at density d. Then with overwhelming probability the following holds.

Let D be a reduced diagram with $|D| = 3$, and suppose D contains a 2-cell corresponding to relator r_1. Then the number of internal 1-cells in D is at most $2d\ell + \varepsilon\ell$.

In particular at $d < 1/4$ the diagram of Figure 12 does not contain the relator r_1.

Proof.
By Theorem 12, the expected number of fulfillings of D is at most

$$\mathbb{E}S_n(D) \leqslant (2m-1)^{\frac{1}{2}(|\partial D| - (1-2d)\ell|D|)} = (2m-1)^{d\ell|D| - L}$$

where $L = \frac{1}{2}(\ell\,|D| - |\partial D|)$ is the internal length of D.

By symmetry of all relators in the presentation, the expected number of fulfillings of D having the fixed relator r_1 as one of its 2-cells is at most $3(2m-1)^{-d\ell}\mathbb{E}S_n(D)$ (the 3 accounts for the choice of the 2-cell in D we are talking about). Thus the probability that there exists such a fulfilling is at most

$$3(2m-1)^{-d\ell}\mathbb{E}S_n(D) \leqslant 3(2m-1)^{2d\ell - L}$$

which decreases exponentially fast if $L \geqslant 2d\ell + \varepsilon\ell$. □

Lemma 45.
Consider a random group at $d < 1/4$. Let c be a 2-cell in $\widetilde{X}$ mapping to r_1. Let Λ_1 be any hypergraph through c. Then, with overwhelming probability, there exists a hypergraph Λ_2 through c which is locally transverse to Λ_1 in the following sense:

If Λ_1 (resp. Λ_2) intersects ∂c at the points x_l, x_r (resp. x_t, x_b), then any 2-cell adjacent to c contains at most one of x_l, x_r, x_t, x_b. In particular, Λ_1, Λ_2 do not form a 2-collared diagram as illustrated in Figure 11.

Moreover, there are at least $(1/2 - 2d - \varepsilon)\ell$ choices for Λ_2.

Proof.
There are two paths p_{top}, p_{bot} from x_l to x_r in ∂c. Define c_{tl} ("topleft") as a 2-cell adjacent to c containing x_l so that the length of the intersection of ∂c_{tl} with p_{top} is maximal (there may be several such 2-cells, just chose one). (Note that thanks to Corollary 11, the intersection of ∂c_{tl} with ∂c is connected.) Let ℓ_{tl} be this length.

Define c_{tr}, c_{bl}, c_{br} and $\ell_{tr}, \ell_{bl}, \ell_{br}$ similarly (see Figure 13).

We have $\ell_{tl} < 2d\ell - \ell_{bl} + \varepsilon\ell$, otherwise the diagram $c \cup c_{tl} \cup c_{bl}$ would contradict Lemma 44. (In order to get a genuine van Kampen diagram in case $c_{tl} \cap c_{tb}$ contains

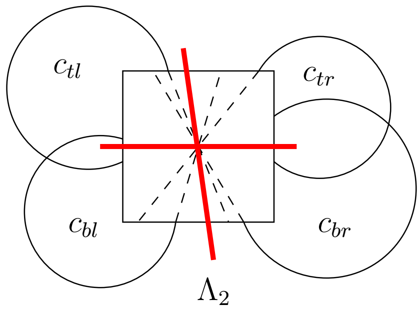


Figure 13:

some 1-cells, we have to unglue a bit c_{tl} below x_l and c_{bl} above x_l — this is consistent with our definition of ℓ_{tl} and ℓ_{bl} as the length of the intersection with resp. p_{top} and p_{bot}).

Similarly, we have $\ell_{tl} \leqslant 2d\ell - \ell_{tr} + \varepsilon\ell$, $\ell_{br} \leqslant 2d\ell - \ell_{bl} + \varepsilon\ell$, and $\ell_{br} \leqslant 2d\ell - \ell_{tr} + \varepsilon\ell$.

Set $L_1 = \max(\ell_{tl}, \ell_{br})$ and $L_2 = \max(\ell_{bl}, \ell_{tr})$. We have $L_1 \leqslant 2d\ell + \varepsilon\ell - L_2$. Since $d < 1/4$ we can choose ε so that $2d + \varepsilon < 1/2$, and so $L_1 < \ell/2 - L_2$ (and the discrepancy is at least $(1/2 - 2d - \varepsilon)\ell$).

Now take any point x_t on p_{top} so that the distance from x_t to x_l lies in the interval $\big(L_1, (\ell/2 - L_2)\big)$. There are at least $(1/2 - 2d - \varepsilon)\ell$ such points. Let $x_b \in p_{bot}$ be the opposite point in c. By construction, x_t and x_b do not lie in any of $c_{tl}, c_{tr}, c_{bl}, c_{br}$. By maximality of these latter 2-cells among 2-cells adjacent to c containing either x_l or x_p, no other 2-cell adjacent to c can contain two of the x's.

Now let of course Λ_2 be the hypergraph through x_t and x_b. □

Lemma 46.
Consider a random group at $d < 1/5$. With overwhelming probability, there is no reduced 2-collared diagram admitting r_1 on one of its corner cells, except the one on Figure 11.

Proof.
Let D be a 2-collared diagram having relator r_1 as one of its corner cells. By Theorem 43, we only have a finite number of diagrams to check. We can thus obtain overwhelming probability by intersecting finitely many events with overwhelming probability.

First, suppose that the other corner 2-cell of D has less than $(1-d)\ell$ edges on the boundary of D. This means that we have $|\partial D| \leqslant \ell + (1-d)\ell + (|D| - 2)\ell/2$. So the expected number of fulfillings of this diagram is, by Theorem 12, at most:

$$\mathbb{E}S_n(D) \leqslant (2m-1)^{\frac{1}{2}(|\partial D| - (1-2d)\ell|D|)} \leqslant (2m-1)^{\ell(1/2 - d/2 + |D|(d-1/4))}$$

By symmetry of all $(2m-1)^{d\ell}$ relators in the presentation, the expected number of fulfillings of D having the fixed relator r_1 as its corner 2-cell is at most $(2m -$

$1)^{-d\ell}\mathbb{E}S_n(D)$, and so the probability that there exists such a fulfilling is at most

$$(2m-1)^{-d\ell}\mathbb{E}S_n(D) \leqslant (2m-1)^{\ell(1/2-3d/2+|D|(d-1/4))}$$

so that if

$$1/2 - 3d/2 + |D|\,(d-1/4) < 0$$

then this probability is exponentially small. So if $|D| > \frac{1/2-3d/2}{1/4-d}$ then with overwhelming probability this does not happen. For $d < 1/5$ the right-hand side is less than 4. So the only possibility is the three 2-cell diagram depicted on Figure 12. But we have just excluded it in Lemma 44.

Second, suppose that the other corner of D has more than $(1-d)\ell$ on the boundary. Then we get the same conclusion by reasoning on the new diagram D' obtained by removing this corner. □

7 Codimension-1 subgroups at $d < 1/5$

Definition 47.
For a hypergraph Λ *in the* 2*-complex* $\widetilde{X}$*, the* orientation-preserving stabilizer $\mathrm{Stabilizer}^+(\Lambda)$ *of* Λ *is the index* $\leqslant 2$ *subgroup of* $\mathrm{Stabilizer}(\Lambda)$ *that also stabilizes each of the two halfspaces which are components of* $\widetilde{X} - \Lambda$*. Equivalently,* $\mathrm{Stabilizer}^+(\Lambda)$ *equals* $\mathrm{Stabilizer}(H^+)$ *where* H^+ *is one of the components of* $\widetilde{X} - \Lambda$*.*

We now prove the existence of codimension-1 subgroups at density $d < 1/5$. These subgroups are orientation-preserving stabilizers of hypergraphs passing through "typical" 2-cells of $\widetilde{X}$.

Lemma 48 (Codimension-1 criterion).
Suppose that the discrete group G *acts cocompactly on the simply connected* 2*-complex* $\widetilde{X}$ *and that the system of hypergraphs in* $\widetilde{X}$ *is locally finite and cocompact. Suppose that the hypergraphs* Λ_1 *and* Λ_2 *cross each other at a single point. Suppose that each* Λ_i *is an embedded tree with no leaves.*

Then $H_i = \mathit{Stabilizer}^+(\Lambda_i)$ *is a subgroup of* G *with relative number of ends* $e(G, H_i) = 2$*.*

Proof.
Let $H_1 = \mathrm{Stabilizer}^+(\Lambda_1)$. Let $\bar{X}_1 = H_1\backslash\widetilde{X}$. We have to prove that $\bar{X}_1$ has at least two ends.

According to Lemma 15, Λ_1 separates $\widetilde{X}$ into two connected components. Let $\bar{\Lambda}_i$ be the image of Λ_i in $\bar{X}_1$. Suppose a component of $\bar{X}_1 - \bar{\Lambda}_1$ is compact. Consider the edge e of Λ_2 where Λ_2 crosses Λ_1. Extend e to a geodesic ray r in the direction of the halfspace mapping to the compact component of $\bar{X}_1 - \bar{\Lambda}_1$.

By compactness, the projection $\bar{r}$ of r to $\bar{X}_1$ must pass through $\bar{\Lambda}_1$ a second time. Indeed, let u be the first combinatorial subpath of r whose projection $\bar{u}$ is a closed path in $\bar{X}_1$ (which exists by compactness). Thus $r = suv$, and s is minimal with this property. Consider the path $p = \bar{s}\bar{u}\bar{s}^{-1}$. We show that p is an immersed path. Indeed, if $\bar{u}\bar{s}^{-1}$ has a backtrack then $\bar{u} = \bar{w}\bar{e}$ and $\bar{s} = \bar{s}'\bar{e}$. Let $\bar{u}' = \bar{e}\bar{w}$. Then $\bar{u}'$ is a closed path with $|\bar{s}'| < |\bar{s}|$, so $\bar{u}'$ occurs earlier than $\bar{u}$ which contradicts the choice of u.

The lift $\tilde{p}$ of p to $\widetilde{X}$ is a segment of Λ_2, which is not closed since it is a subpath of r which is a geodesic in Λ_2.

Finally, let q_1 be a path in Λ_1 which projects to a path in $\bar{\Lambda}_1$ with the same endpoints as p. The common endpoints of p and q_1 provide, after lifting to $\widetilde{X}$, two intersections of the hypergraphs in $\widetilde{X}$, which contradicts the assumption. □

Lemma 49.
With overwhelming probability, at any density, the first relator r_1 in the random presentation involves all generators.

Consequently, hypergraphs have no leaves, and any hypergraph passes through a 2-cell bearing relator r_1.

Proof.
The first assertion is a consequence of the law of large numbers. It follows that any 1-cell of the Cayley 2-complex of a random group is contained in a 2-cell bearing relator r_1; hence, hypergraphs are leafless. □

Theorem 50.
With overwhelming probability, random groups G at density $d < \frac{1}{5}$ have a subgroup H which is free, quasiconvex and such that the relative number of ends $e(G, H)$ is at least 2.

This subgroup can be taken to be the orientation-preserving stabilizer of any hypergraph.

Proof.
Let r_1 be the first relator in the presentation. Let Λ_1 be a hypergraph. By Lemma 49 this hypergraph travels through a 2-cell c bearing r_1. Let Λ_2 be the hypergraph provided by Lemma 45. By Theorem 36, these hypergraphs are embedded trees, leafless by Lemma 49.

By Lemma 46 (in conjunction with Lemma 45), Λ_1 and Λ_2 do not form any reduced collared diagram with corner 2-cell c, and so by Theorem 27 they intersect only at c.

Now apply Lemma 48 to get the number of relative ends. The other assertions follow from Corollary 39. □

Corollary 51.
Suppose that $d < 1/5$. Then with overwhelming probability, a random group does not have Property (T).

Proof.
It was shown in [NR98] that groups having a subgroup with more than one relative end do not have Property (T). □

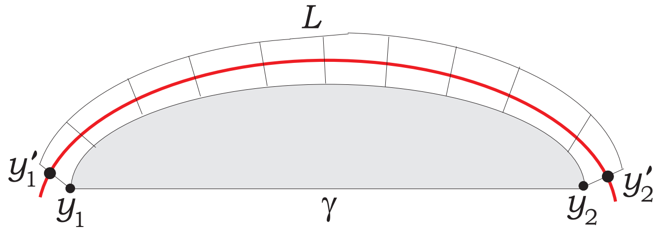


Figure 14:

8 Carriers are convex at $d < 1/6$

Theorem 52.
The following holds with overwhelming probability at $d < \frac{1}{6}$: For each hypergraph Λ its carrier Y is a convex subcomplex of $\widetilde{X}$.

(Recall that the carrier of a hypergraph is the set of 2-cells the hypergraph passes through.)

Proof.
Let y_1, y_2 be two points on Y (which may not lie on the same side of Λ). Let γ be a geodesic in $\widetilde{X}$ joining y_1 to y_2. We want to show that γ lies in Y.

Suppose that γ does not lie in Y. We can decompose γ into subparts which either are included in Y, or intersect Y only at their endpoints. There is nothing to prove in the former case, so we can suppose that the intersection of γ with Y is exactly $\{y_1, y_2\}$ (and in particular y_1 and y_2 lie on the same side of Λ).

Let L be a ladder in Y between 2-cells containing y_1 and y_2, and let y_1' and y_2' be the extremal points of the hypergraph segment contained in L. Let $\gamma' = [y_1'y_1].\gamma.[y_2y_2']$ be the union of γ with paths joining y_i to y_i' respectively. Let D be a reduced van Kampen diagram collared by L and γ', as provided by Lemma 32 (see Figure 14). According to this lemma, D is collared and not only quasicollared since γ' does not intersect the hypergraph except at its endpoints.

Thanks to the collaring, every 2-cell of L except the two extremal ones has less than half its length on the boundary of D.

Now let c be a 2-cell lying on the boundary of D but not belonging to L (so that $\partial c \cap \partial D \subset \gamma$). Since γ is a geodesic, this means that the length of $\partial c \cap \gamma$ is no more than half the boundary length of c (otherwise we could shorten the geodesic).

So every 2-cell of D except maybe the two extremal cells of L has no more than half its length on the boundary of D, so that D has at most two shells. But at $d < 1/6$ this is ruled out by Theorem 40, except when $D = L$ has only two 2-cells, which was to be proven. □

Remark 53.
At density $d < 1/6$, if a 2-cell R is such that ∂R is included in the carrier Y of some hypergraph Λ, then R itself is included in Y. Indeed, otherwise any hypergraph through R would meet Λ twice, contradicting Corollary 41.

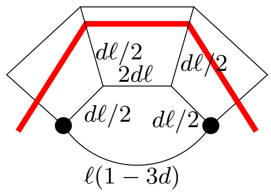


Figure 15: The carrier is not convex at $d > 1/6$.

Remark 54.
It is not difficult to see that carriers are not convex when $d > 1/6$. Indeed (up to some ε's) at density d, with overwhelming probability there exists a diagram as depicted on Figure 15, in which the bottom 2-cell has more than half its boundary length on the boundary of the carrier, thus making it shorter to turn around from below.

9 Separation by hypergraphs at $d < 1/6$

The goal of this section is to prove that for any two points in $\widetilde{X}$, the number of hypergraphs separating them grows linearly with their distance. This will enable us to apply a CAT(0) criterion in Section 10.

For $p, q \in X$ we let $\#(p, q)$ equal the number of hypergraphs Λ such that p and q lie in distinct components of $X - \Lambda$.

Theorem 55.
The following holds with overwhelming probability at $d < \frac{1}{6}$: For all $p, q \in \widetilde{X}^0$ we have:

$$\#(p, q) \geqslant (1/6 - d - \varepsilon)\,(d(p, q) - 3\ell)$$

Corollary 56.
A random group at density $d < 1/6$ has the Haagerup property.

Proof of the corollary.
A discrete group acts properly on its Cayley 2-complex (equipped, say, with the edge metric on the 1-skeleton and Euclidean metrics on each 2-cell). Now, since the hypergraphs embed, the system of hypergraph turns this 2-complex into a space with walls [HP98], and the theorem above states that the wall metric is equivalent to the edge metric. So the group acts properly on a space with walls, which by a folklore remark (see e.g. [CMV04]) implies the Haagerup property. □

For the proof of Theorem 55 we will need two lemmas.

Lemma 57.
The following holds with overwhelming probability at $d < \frac{1}{6}$: For each 2-cell R, the concatenation P_1P_2 of any two disjoint pieces in ∂R satisfies $|P_1|+|P_2| < 3d\ell+\varepsilon\ell < \frac{\ell}{2}$.

Proof.
This follows directly from Theorem 6: indeed, we can form a three-2-cell diagram involving the two pieces, and at $d < 1/6$ its internal length is less than $\ell/2$. □

Lemma 58.
The following holds with overwhelming probability at $d < \frac{1}{6}$: Let Λ be a hypergraph passing through a 2-cell R. Then there exists another hypergraph Λ_d passing through R, such that $\Lambda \cap \Lambda_d$ consists of a single point.

Actually there are at least $(1/2 - 3d - \varepsilon)\ell$ choices for Λ_d.

Proof (Sketch following Lemma 45).
The proof is identical to Lemma 45, except that at density $< 1/6$ we do not have to fix the relator in advance (and we use Lemma 57 in place of Lemma 44). See Figure 13. □

Proof of Theorem 55.
Let γ be a geodesic between p and q. We show that for each length-3ℓ subpath σ of γ there is a hypergraph Λ which crosses an edge of σ but crosses no other edge of γ, and that actually the number of choices for such a Λ in σ is at least $(1/2 - 3d - \varepsilon)\ell$, so that whenever $d(p,q) \geqslant 3\ell$ the number of choices is at least $(1/2 - 3d - \varepsilon)\ell\,(3\ell)^{-1}\,(d(p,q) - 3\ell)$.

Let e_1 be an edge at the middle of σ. Let Λ_e be the corresponding hypergraph. If $\Lambda_e \cap \gamma = \Lambda_e \cap e_1$ then we are done. Otherwise Λ_e crosses γ at a first other edge e_3. Without loss of generality, assume that $e_1 < e_3$ in the ordering on γ.

By convexity, the subpath $[e_1 e_3]$ of γ lies in the carrier of Λ_e.

First observe that e_1 and e_3 cannot be consecutive dual 1-cells of Λ_e crossing the same 2-cell R, for then $|[e_1 e_3]| = \frac{\ell}{2} + 1$ but the complementary part of ∂R has length $\frac{\ell}{2} - 1$ so γ would fail to be a geodesic.

In the other extreme, if there is more than one dual 1-cell of Λ_e between e_1 and e_3, then we let R be the second 2-cell in the ladder of Λ_e between e_1 and e_3. We then apply Lemma 58 to obtain a hypergraph Λ_k passing through R that intersects Λ_e at a single point. This hypergraph crosses γ in a single edge k.

Indeed, suppose Λ_k crossed γ in a second edge k_2. If $e_1 < k_2 < e_3$ then Λ_k crosses Λ_e in a second point which is impossible. Similarly, if $k_2 < e_1$ then by convexity $[k_2 k_1] \subset Y_k$ and hence $e_1 \subset Y_k$. Consequently, Λ_e crosses Λ_k at the center of a 2-cell containing e_1 on its boundary. By hypothesis, this 2-cell cannot be R, and so Λ_e and Λ_k intersect in more than one point which is impossible. An analogous argument excludes the possibility that $e_2 < k_2$.

Finally, we consider the case where e_1 and e_3 are separated by a single dual 1-cell e_2. Let R_e be the 2-cell between e_1 and e_2, and let R be the 2-cell between e_2 and e_3 (see Figure 16).

Applying Lemma 58, let Λ_f be a hypergraph passing through R that intersects Λ_e at a single point. Let Y_f be the carrier of Λ_f.

If Λ_f intersects γ in a single 1-cell f_3 then we are done. Otherwise, let f_1 be the next 1-cell in γ that Λ_f passes through. As in the first case above, since γ is a geodesic,

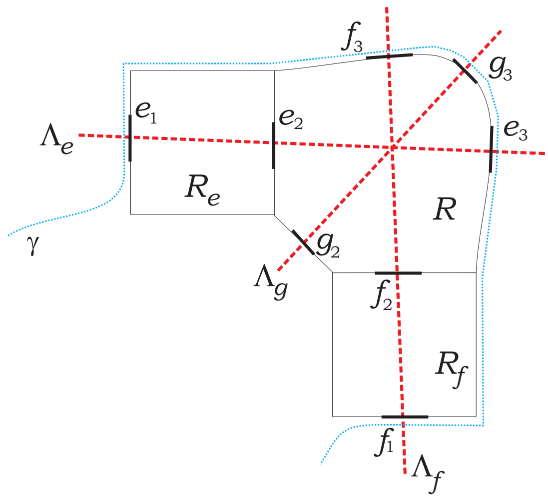


Figure 16:

f_3 and f_1 cannot be consecutive dual 1-cells of Λ_f. Similarly, there cannot be more than one dual 1-cell f_2 between them, for otherwise, as in the second case considered above, we could find a third hypergraph Λ_h intersecting Λ_f at a single point which is at the center of a 2-cell separating f_1 and f_3 and we are done.

Finally, we consider the case where there is exactly one dual 1-cell f_2 between f_3 and f_1 in Y_f. We refer the reader to Figure 16.

Let R_f be the 2-cell in Y_f between f_2 and f_1. Let $P_e = R \cap R_e$ and let $P_f = R \cap R_f$. Note that P_e and P_f are disjoint. Indeed, otherwise the complement of $\gamma \cap R$ is the concatenation of two pieces and this has length $< \frac{\ell}{2}$ by Lemma 57, and so γ would not be a geodesic. Consequently, P_e and P_f are separated on either side of R by $R \cap \gamma$ and an edge g_2.

Let Λ_g be the hypergraph dual to g_2, with carrier Y_g, and let g_3 be the 1-cell in ∂R that is antipodal to g_2. Note that since g_2 lies between e_2 and f_2, the 1-cell g_3 lies between f_3 and e_3 in ∂R, and hence Λ_g crosses γ in g_3. We will show that Λ_g does not cross any other edge of γ.

Note that g_2 is not a 1-cell of γ, since $|[g_2 g_3]| > \frac{\ell}{2}$ and γ is a geodesic.

By definition, g_2 does not lie in $R_e \cap R$. Neither does g_3, since f_3 is between e_2 and g_3 on ∂R and by definition Λ_f does not pass through R_e. So by Corollary 41, Λ_g does not pass through R_e, i.e. $R_e \not\subset Y_g$.

Let a be the last 1-cell in γ before R. Suppose that Λ_g crossed a 1-cell h of γ with $h \leqslant a$. Since R_e does not lie in Y_g we have $h \leqslant e_1$. But then since e_1 lies in the geodesic γ between g_3 and h, by convexity of Y_g we get $e_1 \subset Y_g$. Since $e_2 \subset Y_g$ too by construction, by convexity and Remark 53 we get $R_e \subset Y_g$ which is a contradiction.

Let b be the first 1-cell in γ after R. A similar argument shows that Λ_g cannot cross a 1-cell h of γ with $b \leqslant h$.

Note that we used only three 2-cells in the construction, so that the intersection of γ with these has length at most $3\ell/2$; we have to consider a segment of length 3ℓ

because we do not know on which side of e_1 the edge e_3 will fall. $\square$

10 CAT(0) cubulation at $d < 1/6$

We now proceed to the geometrization theorem at $d < 1/6$. We begin by listing the following criteria from [HW04]:

Theorem 59 (Local Finiteness).
Let $\widetilde{X}$ be a 2–complex equipped with a collection of hypergraphs satisfying the following properties.

1. *$\widetilde{X}$ is locally finite.*
2. *The hypergraph system is uniformly locally finite.*
3. *There is a constant K so that for each $n \geq 1$, every pair of points at a distance at least nK apart are separated by at least n distinct hypergraphs.*
4. *There is a constant δ such that every hypergraph triangle is δ–thin.*

Then the cube complex C associated to $\widetilde{X}$ is locally finite.

Theorem 60 (Properness).
Let $\widetilde{X}$ be locally finite with a locally finite cube complex C. If Γ acts properly discontinuously on $\widetilde{X}$, then the induced action of Γ on C is also properly discontinuous.

The following is formulated in [HW04] but was first proven by Sageev in [Sag97].

Theorem 61 (Cocompactness).
Suppose Γ acts cocompactly on $\widetilde{X}$ then Γ acts cocompactly on C provided that the following conditions hold.

1. *$\widetilde{X}$ is δ–hyperbolic.*
2. *The hypergraphs are quasiconvex.*
3. *The hypergraph system is locally finite.*

We can now prove our second main theorem:

Theorem 62.
With overwhelming probability, a random group at density $d < \frac{1}{6}$ acts freely and cocompactly on a $\mathrm{CAT}(0)$ cube complex C.

Proof.
As shown by Gromov [Gro93] (see also [Oll04]), G is hyperbolic and torsion-free with overwhelming probability at $d < \frac{1}{2}$.

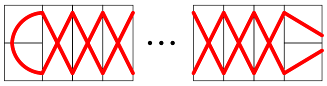


Figure 17: At $d > 1/5$, the hypergraph is 1-dense in $\widetilde{X}$.

The quasiconvexity of the hypergraphs, and hence the codimension-1 subgroups that are their orientation-preserving stabilizers, was proven in Theorem 38 at $d < \frac{1}{5}$ with overwhelming probability. In fact, the convexity of hypergraph carriers was proven in Theorem 52 at $d < \frac{1}{6}$ with overwhelming probability.

The uniform local finiteness of $\widetilde{X}$ and the system of hypergraphs is obvious in our case.

Applying Theorem 61, we see that G acts cocompactly on the cube complex C.

Since G is hyperbolic, and the hypergraphs embed by quasi-isometries, we see that all hypergraph triangles in $\widetilde{X}$ are δ-thin for some δ depending on the hyperbolicity constant for G and the quasi-isometry constants for the hypergraphs.

Finally the linear separation condition was proven in Theorem 55.

Thus the cube complex is locally finite by Theorem 59. Consequently G acts properly discontinuously on C by Theorem 60.

Since G is torsion-free we see that the action is free, and we are done. □

The crucial difference between the 1/5 and 1/6 cases was the separation of any two points by a linear number of hypergraphs proven in Theorem 55. We suspect that this should hold at density $d < 1/5$, but adapting the proof of Theorem 55 to this case involves the analysis of many particular cases corresponding to the existence of small 2-collared diagrams at density $1/6 < d < 1/5$.

Conjecture 63.
With overwhelming probability, random groups at density $d < \frac{1}{5}$ act freely and cocompactly on a CAT(0) cube complex.

11 The unique hypergraph is π_1-surjective at $d > 1/5$

The next theorem shows that our approach fails at density $d > \frac{1}{5}$.

Theorem 64.
Let Λ be a hypergraph in the standard 2-complex $X = \widetilde{X}/G$ of a random group presentation at density $d > \frac{1}{5}$. Then $\pi_1\Lambda \to \pi_1 X \cong G$ is surjective.

Proof.
It is equivalent to prove that in $\widetilde{X}$, if the hypergraph Λ contains the midpoint of edge e_1, then it also contains the midpoint of any other edge e_2 sharing a vertex with e_1. In particular it follows that there is only one hypergraph in $\widetilde{X}$.

We first give a simple argument proving this under the stronger condition that $d > 1/4$. In this case, with overwhelming probability there are two relators sharing a piece of length $\ell/2$. Thus the diagram of Figure 18 occurs with overwhelming

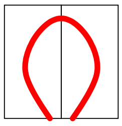


Figure 18: At $d > 1/4$, the hypergraph is trivially 1-dense in $\widetilde{X}$.

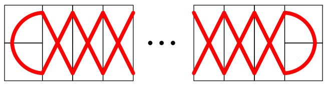


Figure 19:

probability. In fact, there are arbitrarily many such large pieces as $\ell \to \infty$. The distance between the two drawn endpoints of the hypergraph is 1. Even more, this happens arbitrarily many times when $\ell \to \infty$, with overwhelming probability, every combination of generators of the group will appear arise in this way.

We will show that as soon as $d > 1/5$, with overwhelming probability there exists a diagram such as depicted on Figure 17, where the space between the two hypergraph ends on the right is made of two consecutive half-edges bearing arbitrary generators from the presentation.

To show that such a diagram exists, it is enough to show that the diagram in Figure 19 exists. Indeed, the latter is just the former with one less constraint on the relators, and so if the latter exists, then the former does a fortiori. Since there are only a finite number of generators in the presentation, with overwhelming probability all generator combinations on the rightmost "free" edges of the diagram of Figure 17 will occur as $\ell \to \infty$.

First, let us prove that half of this diagram exists, i.e. the diagram in Figure 20. Suppose that $d > 1/5$: then, with overwhelming probability there exist two relators r_1, r_2 sharing a subword of length $2\ell/5$, as on the left of Figure 21. At density d subwords of length $(d-\varepsilon)\ell$ of the relators exhaust all possible words of length $(d-\varepsilon)\ell$, so we can easily find a third relator r_3 gluing to the first two ones along a subword of length $\ell/5$ as in Figure 21. Since $d > 1/5$ we even have on average $(2m-1)^{(d-1/5)\ell}$ choices for r_3. For each of these choices, the rightmost subword w of r_3 (with $|w| = \ell/5$) is chosen at random, so the number of possibilities for w is $(2m-1)^{(d-1/5)\ell}$ (by "number of possibilities" we mean that there are this number of independently picked random words appearing on the right of such a diagram).

We go on constructing the diagram on Figure 20 adding relators on the right side: at each step we have to find a relator gluing to the previous diagram along a subword

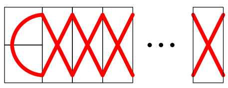


Figure 20:

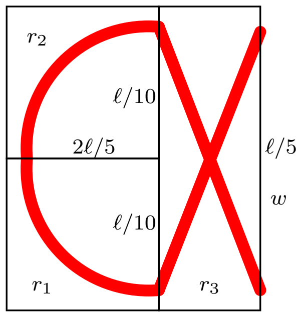


Figure 21: Hypergraphs do not embed at $d > 1/5$.

of length $\ell/5$, which is always possible at $d > 1/5$. Thus, the diagram on Figure 20 exists with overwhelming probability.

We now prove that moreover, for each word w of length $\ell/5$, a diagram like the one on Figure 20 exists with w as the rightmost boundary subword (for a large enough number of intermediary cells). Indeed, we have seen that the number of choices in the preceding construction is multiplied by $(2m-1)^{(d-1/5)\ell}$ at each step. So after $1/(d-1/5)$ steps, the number of choices (i.e. independent pickings) for the rightmost $\ell/5$-long subword will be $(2m-1)^{\ell}$, much bigger than the total number of words of this length, so that with overwhelming probability all possible words will be picked.

Now since we can choose the rightmost word on Figure 20, we can pick two such diagrams with inverse rightmost words and glue them to get the diagram on Figure 19. The number of necessary intermediate cells behaves roughly like $1/(d-1/5)$. □

References

[BdlHV08] Bachir Bekka, Pierre de la Harpe, and Alain Valette. *Kazhdan's property (T)*, volume 11 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2008.

[BJS88] M. Bożejko, T. Januszkiewicz, and R. J. Spatzier. Infinite Coxeter groups do not have Kazhdan's property. *J. Operator Theory*, 19(1):63–67, 1988.

[BŚ97] W. Ballmann and J. Świątkowski. On L^2-cohomology and property (T) for automorphism groups of polyhedral cell complexes. *Geom. Funct. Anal.*, 7(4):615–645, 1997.

[CCJ+01] Pierre-Alain Cherix, Michael Cowling, Paul Jolissaint, Pierre Julg, and Alain Valette. *Groups with the Haagerup property*, volume 197 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 2001. Gromov's a-T-menability.

[CDH] Indira Chatterji, Cornelia Druţu, and Frédéric Haglund. Median spaces and spaces with measured walls. Preprint.

[CMV04] Pierre-Alain Cherix, Florian Martin, and Alain Valette. Spaces with measured walls, the Haagerup property and property (T). *Ergodic Theory Dynam. Systems*, 24(6):1895–1908, 2004.

[CN04] I.L. Chatterji and G.A. Niblo. From wall spaces to CAT(0) cube complexes. 2004.

[dlHV89] Pierre de la Harpe and Alain Valette. *La propriété (T) de Kazhdan pour les groupes localement compacts*. Number 175 in Astérisque. Soc. Math. Fr, 1989. With an appendix by M. Burger.

[Ghy04] Étienne Ghys. Groupes aléatoires (d'après Misha Gromov,...). *Astérisque*, (294):173–204, 2004. Séminaire Bourbaki, Vol. 2003/2004, Exp. 916.

[Gro93] M. Gromov. Asymptotic invariants of infinite groups. In *Geometric group theory, Vol. 2 (Sussex, 1991)*, pages 1–295. Cambridge Univ. Press, Cambridge, 1993.

[Hou74] C. H. Houghton. Ends of locally compact groups and their coset spaces. *J. Austral. Math. Soc.*, 17:274–284, 1974. Collection of articles dedicated to the memory of Hanna Neumann, VII.

[HP98] Frédéric Haglund and Frédéric Paulin. Simplicité de groupes d'automorphismes d'espaces à courbure négative. In *The Epstein birthday schrift*, pages 181–248 (electronic). Geom. Topol., Coventry, 1998.

[HW04] Chris Hruska and Daniel T. Wise. Axioms for finiteness of cubulations. Preprint, 2004.

[LS77] Roger C. Lyndon and Paul E. Schupp. *Combinatorial group theory*. Springer-Verlag, Berlin, 1977. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 89.

[MW02] Jonathan P. McCammond and Daniel T. Wise. Fans and ladders in small cancellation theory. *Proc. London Math. Soc. (3)*, 84(3):599–644, 2002.

[Nic04] Bogdan Nica. Cubulating spaces with walls. *Algebr. Geom. Topol.*, 4:297–309 (electronic), 2004.

[NR97] Graham Niblo and Lawrence Reeves. Groups acting on CAT(0) cube complexes. *Geom. Topol.*, 1:approx. 7 pp. (electronic), 1997.

[NR98] Graham A. Niblo and Martin A. Roller. Groups acting on cubes and Kazhdan's property (T). *Proc. Amer. Math. Soc.*, 126(3):693–699, 1998.

[NR03] G. A. Niblo and L. D. Reeves. Coxeter groups act on CAT(0) cube complexes. *J. Group Theory*, 6(3):399–413, 2003.

[Oll04] Y. Ollivier. Sharp phase transition theorems for hyperbolicity of random groups. *GAFA, Geom. Funct. Anal.*, 14(3):595–679, 2004.

[Oll05a] Yann Ollivier. Cogrowth and spectral gap of generic groups. *Ann. Inst. Fourier (Grenoble)*, 55(1):289–317, 2005.

[Oll05b] Yann Ollivier. *A January 2005 invitation to random groups*, volume 10 of *Ensaios Matemáticos [Mathematical Surveys]*. Sociedade Brasileira de Matemática, Rio de Janeiro, 2005.

[Oll07] Yann Ollivier. Some small cancellation properties of random groups. *Internat. J. Algebra Comput.*, 17(1):37–51, 2007.

[Sag95] Michah Sageev. Ends of group pairs and non-positively curved cube complexes. *Proc. London Math. Soc. (3)*, 71(3):585–617, 1995.

[Sag97] Michah Sageev. Codimension-1 subgroups and splittings of groups. *J. Algebra*, 189(2):377–389, 1997.

[Sco78] Peter Scott. Ends of pairs of groups. *J. Pure Appl. Algebra*, 11(1-3):179–198, 1977/78.

[Ser80] Jean-Pierre Serre. *Trees*. Springer-Verlag, Berlin, 1980. Translated from the French by John Stillwell.

[Wis04] Daniel T. Wise. Cubulating small cancellation groups. *GAFA, Geom. Funct. Anal.*, 14(1):150–214, 2004.

[Żuk03] A. Żuk. Property (T) and Kazhdan constants for discrete groups. *Geom. Funct. Anal.*, 13(3):643–670, 2003.

On a small cancellation theorem of Gromov

Ce texte date de 2004, pendant ma thèse. Il est reproduit ici car il sert de base à l'article ultérieur Kazhdan groups with infinite outer automorphism group *avec D. Wise. Il a été publié au volume 13, n°1 (2006) du* Bulletin of the Belgian Mathematical Society – Simon Stevin*, pp. 75–89.*

On a small cancellation theorem of Gromov

Yann Ollivier

Abstract

We give a combinatorial proof of a theorem of Gromov, which extends the scope of small cancellation theory to group presentations arising from labelled graphs.

In this paper we present a combinatorial proof of a small cancellation theorem stated by M. Gromov in [Gro03], which strongly generalizes the usual tool of small cancellation. Our aim is to complete the six-line-long proof given in [Gro03] (which invokes geometric arguments).

Small cancellation theory is an easy-to-apply tool of combinatorial group theory (see [Sch73] for an old but nicely written introduction, or [GH90] and [LS77]). In one of its forms, it basically asserts that if we face a group presentation in which no two relators share a common subword of length greater than 1/6 of their length, then the group so defined is hyperbolic (in the sense of [Gro87], see also [GH90] or [Sho91] for basic properties), and infinite except for some trivial cases.

The theorem extends these conclusions to much more general situations. Suppose that we are given a finite graph whose edges are labelled by generators of the free group F_m and their inverses (in a reduced way, see definition below). If no word of length greater than 1/6 times the length of the smallest loop of the graph appears twice on the graph, then the presentation obtained by taking as relations all the words read on all loops of the graph defines a hyperbolic group which (if the rank of the graph is at least $m+1$, to avoid trivial cases) is infinite. Moreover, the given graph naturally embeds isometrically into the Cayley graph of the group.

The new theorem reduces to the classical one when the graph is a disjoint union of circles. Noticeably, this criterion is as easy to use as the standard one.

For example, ordinary small cancellation theory cannot deal with such simple group presentations as $\langle\, S \mid w_1 = w_2 = w_3 \,\rangle$ because the two relators involved here, $w_1 w_2^{-1}$ and $w_1 w_3^{-1}$, share a long common subword. The new theorem can handle such situations: for "arbitrary enough" words w_1, w_2, w_3, such presentations will define infinite, hyperbolic groups, although from the classical point of view these presentations satisfy (e.g. if the w_i's have the same length) a priori only the $C'(1/2)$ condition from which nothing could be deduced.

The groups obtained by this process can in some cases be noticeably different from ordinary small cancellation groups. For example, the graphs used by Gromov in [Gro03] provide groups having Kazhdan's property (T) (see [Sil03]), whereas ordinary small cancellation groups cannot have property (T) (see [Wis04]).

Most importantly, this technique allows to (quasi-)embed prescribed graphs into the Cayley graphs of hyperbolic groups. It is the basic construction involved in the announcement of a counter-example to the Baum-Connes conjecture with coefficients (see [HLS02] which elaborates on [Gro03], or [Ghy03] for a survey). Indeed, this counter-example is obtained by constructing a finitely generated group (which is a limit of hyperbolic groups) whose Cayley graph quasi-isometrically contains an infinite family of expanders.

Moreover, this technique will be used in [OW] to construct new examples of groups with property (T).

1 Statement and discussion

Let S be a finite set, in which an involution without fixed point, called *being inverse*, is given. The elements of S are called *letters*.

A *word* is a finite sequence of letters. The inverse of a word is the word made of the inverse letters put in reverse order. A word is called *reduced* if it does not contain a letter immediately followed by its inverse.

A *labelled graph* is an unoriented graph in which each unoriented edge is considered as a couple of two oriented edges, and each oriented edge bears a letter such that opposite edges bear inverse letters. We require maps of labelled graphs to preserve the labels.

A labelled graph is said to be *reduced* if there is no pair of oriented edges arising from the same vertex and bearing the same letter.

Note that a word can be seen as a (linear) labelled graph, which we will implicitly do from now on. The word is reduced if and only if the labelled graph is.

A *piece* of a labelled graph is a word which has two different immersions in the labelled graph. (An immersion is a locally injective map of labelled graphs. Two immersions are considered different if they are different as maps.) This is analogue to the traditional piece of small cancellation theory.

A *standard family of cycles* for a connected graph is a set of paths in the graph, generating the fundamental group, such that there exists a maximal subtree of the graph such that, when the subtree is contracted to a point (so that the graph becomes a bouquet of circles), the set of generating cycles is exactly the set of these circles. There always exists some. If the graph is not connected, a standard family of cycles is one which is standard on each component.

A *generating family of cycles* is a family of cycles generating the fundamental group of each connected component of the graph (maybe up to adding initial and final segments joining these cycles to some basepoint).

A graph is *non-filamenteous* if every edge belongs to some immersed cycle.

We are now in a position to state the theorem.

Theorem 1 (M. Gromov, [Gro03]).
Let Γ be a finite reduced non-filamenteous labelled graph. Let R be the set of words read on all cycles of Γ (or on a generating family of cycles). Let g be the girth of Γ and Λ be the length of the longest piece of Γ.

If $\Lambda < g/6$ then the presentation $\langle\, S \mid R \,\rangle$ defines a group G enjoying the following properties.

1. *It is hyperbolic, torsion-free.*
2. *Any presentation of G by the words read on a standard family of cycles of Γ is aspherical (in the sense of Definition 9), hence the cohomological dimension of G is at most 2.*
3. *The Euler characteristic of G is $\chi(G) = 1 - |S|/2 + b_1(\Gamma)$. In particular, if the rank of the fundamental group of Γ is greater than the number of generators, G is infinite and not quasi-isometric to $\mathbb{Z}$.*
4. *The shortest relation in G is of length g.*
5. *For any reduced word w representing the identity in G, some cyclic permutation of w contains a subword of a word read on a circle immersed in Γ, of length at least $(1 - 3\Lambda/g)$ (which is more than $1/2$) times the length of this cycle.*
6. *The natural maps from each connected component of the labelled graph Γ into the Cayley graph of G are isometric embeddings.*

If Γ is a disjoint union of circles, this theorem almost reduces to ordinary $1/6$ small cancellation theory. The "almost" accounts for the fact that the length of a shared piece between two relators is supposed to be less than $1/6$ the length of the smallest of the two relators in ordinary small cancellation theory, and less than $1/6$ the length of the smallest of all relators in our case; this is handled through the following remark (which we do not prove in order not to have still heavier notation).

Remark 2.
It is clear from the proof that the assumption in the theorem can be replaced by the following slightly weaker one: for each piece, its length is less than $1/6$ times the length of any cycle of the graph on which the piece appears.

With this latter assumption, the theorem reduces to ordinary small cancellation when the graph is a disjoint union of circles.

Remark 3.
Non-filamenteousness is needed only to ensure isometric embedding of the graph (filaments may not embed isometrically if $\Lambda \geqslant g/8$).

The group obtained is not always non-elementary: for example, if there are three generators a, b, c and the graph consists in two points joined by three edges bearing a, b and c respectively, one obtains the presentation $\langle\, a, b, c \mid a = b = c \,\rangle$ which defines $\mathbb{Z}$.

However, since the cohomological dimension is at most 2, it is easy to check (computing the Euler characteristic) that if the rank of the fundamental group of Γ is greater than the number of generators, then G is non-elementary.

This theorem is not stated explicitly in [Gro03] in the form we give but using a much more abstract and more powerful formalism of "rotation families of groups" ([Gro03], section 2). In the vocabulary thereof, the case presented here is when this rotation family contains only one subgroup of the free group (and its conjugates), namely the one generated by the words read on cycles of the graph with some base point; the corresponding "invariant line" U is the universal cover of the labelled graph Γ (viewed embedded in the Cayley graph of the free group). Reducedness of the labelling ensures convexity.

Elements for a proof of the theorem for very small values of Λ/g (instead of $\Lambda/g < 1/6$) using geometric rather than combinatorial tools, can be found in [Gro01] (see also [Gro03], p. 88).

In [Gro03], this theorem is applied to a random labelling (or rather a variant, Theorem 18 below, in which reducedness is replaced with quasi-geodesicity). It is not difficult, using for example the techniques described in [Oll04], to check that a random labelling satisfies the small cancellation and quasi-geodesicity assumptions.

Acknowledgements. I would like to thank Thomas Delzant for having brought the problem to my attention and for very careful reading, and Étienne Ghys, Pierre Pansu and Alain Valette for helpful discussions and comments on the manuscript.

2 Idea of proof

The line of the argument is as follows: Choose a presentation of G by the words read on a standard generating family of cycles of Γ. We will study the isoperimetry of van Kampen diagrams with respect to this set of relations: we will show that the number of faces in such diagrams is linearly bounded by its boundary length.

Define a labelled complex Γ_2 by attaching to Γ a disk for each cycle in the family. Now each face of a van Kampen diagram for this presentation can be lifted (in a unique way) to Γ_2. For any edge between two faces of the diagram, either these two faces are already adjacent along "the same" edge in Γ_2 or they are not.

Decompose the diagram into maximal parts all edges of which originate from Γ_2 in this sense. Now gluings between these parts do not originate from Γ_2 and thus constitute pieces. So these parts are in classical $1/6$ small cancellation with respect to each other, and so the boundary length of the diagram is controlled in terms of the boundary lengths of these parts. We get the other usual consequences of small cancellation theory as well (asphericity, radius of injectivity...). Technicalities arise from the necessity to perform some so-called "diamond moves" and from the maybe non-simple connectedness of these parts.

To reach the conclusion it is then enough to work inside each part. Since each part lifts to Γ_2 its boundary word is the word read on some null-homotopic cycle in

Γ_2. So this cycle is the product of elements our generating family of cycles, and for isoperimetry we have to control the number of terms in this product (the number of faces in the part) in function of the length of the cycle (the boundary length of the part). This is achieved by decomposing the considered cycle into a product of cycles shorter than three times the diameter of the graph. As there are only finitely many such short cycles we are done.

3 Proof (expanded version)

We now give some more definitions which are useful for the proof.

Definition 4.
A labelled complex *is a finite unoriented combinatorial* 2*-complex the interior of every face of which is homeomorphic to an open disk D_{n+1} with $n \geqslant 0$ holes (n depends on the face), such that its* 1*-skeleton is equipped with a labelled graph structure.*

A labelled complex is said to be reduced *if its* 1*-skeleton is.*

Each face of such a complex defines a set of *contour words*: If the interior of the face is homeomorphic to an open disk D_{n+1} with n holes, the contour words are the $n+1$ cyclic words read by moving around the $n+1$ boundary components of D_{n+1}. The words in this set are considered as oriented cyclic words, and counted with multiplicities.

We require a map of labelled complexes to preserve labels (but it may change orientation of faces, sending a face to a face with inverse contour labels — this amounts to considering maps between the corresponding oriented complexes).

Definition 5.
A tile *is a planar labelled complex with only one face (not necessarily simply connected) and each edge of which belongs to the combinatorial boundary of the face with multiplicity one. We do not fix the embedding in the plane.*

It follows from the definition that the contour of a tile coincides with its boundary.

By our definition of maps between labelled complexes, a tile is considered equal to the tile bearing the inverse boundary words.

Convention: A tile may bear a word which is not simple (i.e. is a power of a smaller word). In this case the tile would have a non-trivial automorphism. To prevent this, say that on each boundary component of a tile we mark a starting point and that a map between tiles has to preserve marked points. This is useful for the study of asphericity and torsion (see Definition 9).

To any planar labelled complex with only one face we can associate a tile in the following way: First, remove the edges that do not belong to the adherence of the interior of the complex (the "filaments"). Then, the obtained one-face complex immersed in the plane is the image of some one-face complex embedded in the plane by a cellular map (this complex is constructed by ungluing along the internal edges). This is an embedding in the plane of some tile, which we call the *tile associated to* the one-face labelled complex.

Definition 6.
A tile of a labelled complex *is the tile associated to any of its faces.*

The *length* of a tile is the length of its boundary.

Definition 7.
A piece *with respect to a set of tiles is a word which has immersions in the boundary of two different tiles, or two distinct immersions in the boundary of one tile.*

Definition 8.
A puzzle *with respect to a set of tiles is a planar labelled complex all tiles of which belong to this set of tiles (the same tile may appear several times in a puzzle). The* set of boundary words *of a puzzle is the set of words read on its boundary components (with multiplicities and orientations).*

A spherical puzzle *is the same drawn on a sphere instead of the plane, that is, a labelled complex which is a combinatorial 2-sphere, all tiles of which belong to this set of tiles.*

A puzzle is said to be minimal *if it has the minimal number of tiles among all puzzles having the same set of boundary words.*

A puzzle is said to be van Kampen-reduced *if there is no pair of adjacent faces such that the words read on the external contour of these two faces are inverse and the position (with respect to the marked point) of the letter read at a common edge of these faces is the same in the two copies of the contour word of these faces.*

So a puzzle is roughly speaking a van Kampen diagram in which we allow non-simply connected faces. The last definition given corresponds to reduced van Kampen diagrams (see [LS77]). (Incidentally, a reduced puzzle is van Kampen-reduced, though the converse is not necessarily true.)

Definition 9.
A presentation of a group is said to be aspherical *if the set of tiles whose boundary words are the relators of the presentation admits no van Kampen-reduced spherical puzzle.*

There are several notions of aspherical presentations in the literature (see e.g. [CCH81] for five of them). Our definition of asphericity coincides with the one in [Ger87], p. 31 (in which asphericity is termed "every spherical diagram is diagrammatically reducible"). It is thus stronger than the one(s) in [LS77], the main difference being that we mark a starting point on the boundary of each tile (see the discussion in [Ger87]). In particular, with our (and [Ger87]'s, contra [LS77]) convention, a presentation such as $\langle\, S \mid w^n = 1 \,\rangle$ (with $n \geqslant 2$) is not aspherical: no relator can be a proper power. With this convention, asphericity of a presentation implies asphericity of the Cayley 2-complex ([Ger87], p. 32), hence (by Hurewicz' Theorem) cohomological dimension at most 2 and hence ([Bro82], p. 187) torsion-freeness.

Proof of the theorem.
Let Γ be a reduced labelled graph. The group under consideration is defined by the

presentation $\langle\, S \mid R \,\rangle$ where R is the set of all words read along cycles of Γ. However, taking all words is not necessary: the group presented by $\langle\, S \mid R \,\rangle$ will be the same if we take not all cycles but only a generating set of cycles.

The fundamental group of the graph Γ is a free group. Let $\mathcal{C}$ be a finite generating set of $\pi_1(\Gamma)$ (maybe not standard). Let R be the set of words read on the cycles in $\mathcal{C}$.

Add 2-faces to Γ in the following way: for each cycle in $\mathcal{C}$, glue a disk bordering this cycle. Denote by Γ_2 this 2-complex; it depends on the choice of $\mathcal{C}$, or equivalently on R.

As the cycles in $\mathcal{C}$ generate all cycles, Γ_2 is simply connected. Note that if $\mathcal{C}$ happens to be taken standard, as will sometimes be the case below, then Γ_2 has no homotopy in degree 2.

By our definitions above (Definition 6), a tile of Γ_2 is a topological disk whose boundary is labelled by some word of R.

We are going to show that there exists a constant $C > 0$ such that any simply connected van Kampen-reduced puzzle D with respect to the tiles of Γ_2 satisfies a linear isoperimetric inequality $|\partial D| \geqslant C\,|D|$ where $|\partial D|$ is the boundary length of D and $|D|$ is the number of faces of D. This implies hyperbolicity (see for example [Sho91]).

We can safely assume that all edges of D lie on the contour of some face (roughly speaking, there are no "filaments"). Indeed, filaments only improve isoperimetry. Generally speaking, in what follows we will never mention the possible occurrence of filaments, their treatment being immediate.

Remark 10.
The 1/6 assumption on pieces implies that no two distinct cycles of Γ bear the same word.

Let e be an internal[1] edge of D, adjacent[2] to faces f_1 and f_2. As D is a puzzle over the tiles of Γ_2, there are faces f'_1 and f'_2 of Γ_2 bearing the same contour words as f_1 and f_2 respectively (maybe up to inversion). These faces are unique by Remark 10.

The edge e belongs to the contour of both f_1 and f_2 and thus can be lifted in Γ_2 either in f'_1 or in f'_2. Say e is an *edge originating from* Γ_2 if these two lifts coincide, so that in Γ_2, the two faces at play are adjacent along the same edge as they are in D.

Any labelled complex with respect to the tiles of Γ_2, all internal edges of which originate from Γ_2, can thus be lifted to Γ_2 by lifting each of its edges. This lifting is unique by Remark 10.

Note that D is van Kampen-reduced if and only if there is no edge e originating from Γ_2 and adjacent to faces f_1, f_2 such that $f'_1 = f'_2$.

We work by first proving the isoperimetric inequality for puzzles having all edges originating from Γ_2. Second, we will decompose the puzzle D into "parts" having all

[1] i.e. not on the boundary

[2] We say that two faces f_1, f_2 of a 2-complex are *adjacent* along edge e (or simply *adjacent* if the mention of e is unnecessary) if either $f_1 \neq f_2$ and e belongs to the contour of both f_1 and f_2, or $f_1 = f_2$ and e is included twice in the contour of f_1.

their edges originating from Γ_2 and show that these parts are in $1/6$ small cancellation with each other. Then we will use ordinary small cancellation theory to conclude.

We begin by proving what we want for some particular choice of R.

Lemma 11.
Let $\Delta = \mathrm{diam}(\Gamma)$. *Suppose that* $\mathcal{C}$ *was chosen to be the set of closed paths embedded (or immersed) in* Γ *of length at most* 3Δ. *Then, for any closed path in* Γ *labelling a reduced word* w, *there exists a simply connected puzzle with boundary word* w, *with tiles having their boundary words in* R, *all edges of which originate from* Γ_2, *and with at most* $3\,|w|\,/g$ *tiles.*

Proof of Lemma 11.
If $|w| \leqslant 2\Delta$ then by definition of R there exists a one-tile puzzle spanning w, and as $|w| \geqslant g$ the conclusion holds. Show by induction on n that if $|w| \leqslant n\Delta$ there exists a simply connected puzzle D spanning w with at most n tiles. This is true for $n = 2$. Suppose this is true up to $n\Delta$ and suppose that $2\Delta \leqslant |w| \leqslant (n+1)\Delta$.

Let $w = w'w''$ where $|w'| = 2\Delta$. As the diameter of Γ is Δ, there exists a path in Γ labelling a word x joining the endpoints of w', with $|x| \leqslant \Delta$. So $w'x^{-1}$ is read on a cycle of Γ of length at most 3Δ, hence (its reduction) belongs to R. Now xw'' is a word read on a cycle of Γ, of length at most $|w| - \Delta \leqslant n\Delta$. So there is a puzzle with at most n tiles spanning xw''. Gluing this puzzle with the tile spanning $w'x^{-1}$ along the x-sides provides the desired puzzle. (Note that this gluing occurs in Γ_2, so that edges of the resulting puzzle originate from Γ_2.)

So for any w we can find a puzzle spanning it with at most $1 + |w|\,/\Delta$ tiles. As $\Delta \geqslant g/2$ and as $|w| \geqslant g$, we have $1 + |w|\,/\Delta \leqslant 1 + 2\,|w|\,/g \leqslant 3\,|w|\,/g$. □

Corollary 12.
For any choice of $\mathcal{C}$, *there exists a constant* $\alpha > 0$ *such that any minimal simply connected puzzle* D *with respect to the tiles of* Γ_2 *all internal edges of which originate from* Γ_2 *satisfies the isoperimetric inequality* $|\partial D| \geqslant \alpha\,|D|$.

Proof of Corollary 12.
Indeed, the existence of an isoperimetric constant for minimal diagrams does not depend on the finite presentation, hence the result when $\mathcal{C}$ is finite. This also holds for infinite $\mathcal{C}$ since any infinite family of cycles in the finite graph Γ contains a finite generating subfamily. □

These last affirmations only express in terms of diagrams the fact that the fundamental group of Γ, which is free hence hyperbolic, is generated by the cycles of Γ of length at most 3Δ (w.r.t. some basepoint).

The next lemma is just ordinary small cancellation theory (see for example the appendix of [GH90], or [LS77]), stated in the form we need. Note that usually, the definition of small cancellation involves pieces of relative size less than λ with $\lambda \leqslant 1/6$. Here we use pieces of relative size at most λ with $\lambda < 1/6$. This is less well-suited for treatment of infinite presentations (which we do not consider) but allows lighter notation for the isoperimetric constant $1-6\lambda > 0$ and the Greendlinger constant $1-3\lambda > 1/2$.

Lemma 13.
Let R be a set of simply connected reduced tiles. Suppose that any piece with respect to two tiles $t, t' \in R$ is a word of length at most λ times the smallest boundary length of t and t', for some constant $\lambda < 1/6$.

Then any simply connected van Kampen-reduced puzzle D with respect to the tiles of R satisfies the following properties.

1. *If D has at least two faces, the reduction w of the boundary word of D contains two disjoint subwords w_1, w_2, with w_1 (resp. w_2) subword of the boundary word of some tile t_1 (resp. t_2) of D, with length at least $(1-3\lambda) > \frac{1}{2}$ times the boundary length of t_1 (resp. t_2).*

2. *The word w is not a proper subword of the boundary word of some tile.*

3. *The boundary length $|\partial D|$ is at least $1-6\lambda$ times the sum of the lengths of the faces of D, and at least the boundary length of the largest tile it contains.*

Moreover, there is no spherical van Kampen-reduced puzzle with respect to these tiles.

Corollary 14.
Let R be a set of (not necessarily simply connected) reduced tiles. Suppose that any piece with respect to two tiles $t, t' \in R$ is a word of length at most λ times the smallest length of the boundary component of t and t' it immerses in, for some constant $\lambda < 1/6$.

Then, any simply connected puzzle with respect to this set of tiles contains only simply connected tiles.

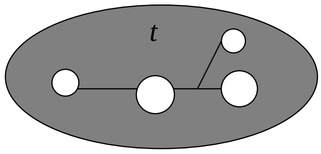
Proof of the corollary.

Let D be a simply connected puzzle with respect to R. Let t be a non-simply connected tile in D. We can suppose that t is deepest, that is, that the bounded components of the complement of t contain no other non-simply connected tile.

The interior of t is embedded in the plane and is homeomorphic to a disk with some finite number n of holes. Since D is simply connected any such hole is filled with a subpuzzle. So let $D'_1, \ldots, D'_n$ be the subpuzzles filling the bounded connected components of the complement of the interior of t. Each D'_i is simply connected, since the bounded connected components of the complement of a connected set in the plane are simply connected. Let us work with D'_1. In case D'_1 is not van Kampen-reduced we replace it by its van Kampen-reduction (which does not change its boundary word, so it can still be glued to one of the holes of t).

The boundary of D'_1 may not be embedded in the plane. However, it is immersed, since the word read on it is the word read on one of the interior boundaries of t, and this word is reduced.

The component D'_1 is a connected simply connected puzzle. Its image in the plane is the union of closed sets $D''_1, \dots, D''_q$ such that each D''_i is either a topological closed disk or a topological closed segment ("filament"), and the D''_i's intersect at a finite number of points. By construction, each D''_i which is a disk is a puzzle.



Suppose that D''_i is a segment. Then each of its endpoints belongs to some D''_j with $j \neq i$. Indeed, otherwise the boundary of D'_1 would not be immersed.

Construct a graph T embedded in the plane in the following way. For each D''_i which is a disk, define a family of segments T_i as follows: Choose a point p_0 in the interior of D''_i. There are a finite number of points $p_1, \dots, p_r$ on the boundary of D''_i such that p_j belongs to some D''_k for $k \neq i$. Now define T_i to be made of the union of segments $p_0p_j \subset D''_i$ for $1 \leqslant j \leqslant r$. Now define T to be the union of all D''_i for those $1 \leqslant i \leqslant q$ for which D''_i is a segment, plus the union of all T_i's for those $1 \leqslant i \leqslant q$ for which D''_i is a disk.

By construction, T is connected since D'_1 is.

For each i such that D''_i is a disk, D''_i retracts onto T_i preserving the points $p_1, \dots, p_r$. So D'_1 retracts onto T, and in particular T is simply connected since D'_1 is. So T is a tree. It is non-empty since D'_1 is (but maybe reduced to a point if D'_1 is a topological disk).

Now consider some leaf of T. Since any endpoint of any D''_i which is a segment belongs to some D''_j with $j \neq i$ (since $\partial D'_1$ is immersed as we saw above), a leaf of T cannot belong to a D''_i which is a segment. So a leaf of T belongs to some T_i constructed from some D''_i which is a disk. By definition of T_i, this means that D''_i intersects with at most one other D''_j with $j \neq i$.

Now D''_i is a puzzle which is a topological disk. As we supposed that t was taken a deepest non-simply connected tile, D''_i contains only simply connected tiles. So we can apply Lemma 13: there exist two tiles t', t'' in D''_i and two subwords w', w'' of the boundary word of D''_i such that w' (resp. w'') is a subword of the boundary word of t' (resp. t'') of length at least one half the boundary length of t' (resp. t''). As D''_i has at most one point of intersection with the other D''_j for $j \neq i$, at least one of w' and w'' is a subword of the boundary of D'_1. But a boundary word of D'_1 is a boundary word of the tile t, and so t shares with t' or t'' a word of length at least one half the boundary length of t' or t'', which contradicts the small cancellation assumption. □

Back to our simply connected van Kampen-reduced minimal puzzle D with tiles in Γ_2. A puzzle is built by taking the disjoint union of all its tiles and gluing them along the internal edges.

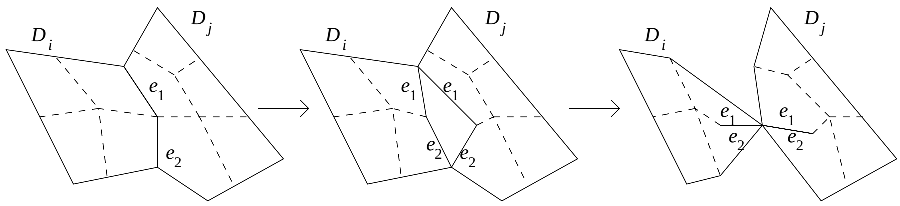
First, define a disjoint union of puzzles D' by taking the disjoint union of all tiles of D and gluing them along the internal edges of D originating from Γ_2. All internal edges of D' originate from Γ_2.

As D is van Kampen-reduced, D' is as well.

Let D_i, $i = 1, \ldots, n$ be the connected components of D'. They form a partition of D. The puzzle D is obtained by gluing these components along the internal edges of D not originating from Γ_2.

It may be the case that the boundary word of some D_i is not reduced. This means that there is a vertex on the boundary of D_i which is the origin of two (oriented) edges bearing the same vertex. We will modify D in order to avoid this. Suppose some D_i has non-reduced boundary word and consider two edges e_1, e_2 of D responsible for this: e_1 and e_2 are two consecutive edges with inverse labels. These edges are either boundary edges of D or internal edges. In the latter case this means that D_i is to be glued to some D_j. We treat only this latter case as the other one is even simpler.

Make the following transformation of D: do not glue any more edge e_1 of D_i with edge e_1 of D_j, neither edge e_2 of D_i with edge e_2 of D_j, but rather glue edges e_1 and e_2 of D_i, as well as edges e_1 and e_2 of D_j, as in the following picture. This is possible since by definition e_1 and e_2 bear inverse labels.



This kind of operation has been studied and termed *diamond move* in [CH82]. The case when the central point has valency greater than 2 (i.e. when more than two D_i's meet at this point) is treated similarly.

Since Γ_2 is reduced, the lifts to Γ_2 of the edges e_1 and e_2 of D_i are the same edge of Γ_2. This shows that the transformation above preserves the fact that all edges of D_i and of D_j originate from Γ_2.

The resulting puzzle (denoted D again) has the same number of faces as before, and no more boundary edges. Thus, proving isoperimetry for the modified puzzle will imply isoperimetry for the original one as well. So we can safely assume that the boundary words of the D_i's are reduced.

Now consider D as a puzzle with the D_i's as tiles. (More precisely, if we erase from D all internal edges originating from Γ_2 then we obtain a puzzle each tile of which is the tile associated to the one-face complex obtained from some D_i by erasing all internal edges originating from Γ_2.) This is a van Kampen-reduced puzzle, since if D_i and D_j are in reduction position this means that they lift to the same subcomplex of Γ_2 and share an edge originating from Γ_2, which contradicts their definition. Note that these tiles are not necessarily simply connected.

These tiles satisfy the condition of Corollary 14. Indeed, suppose that two tiles D_i, D_j (with maybe $i = j$ in which case two parts of the boundary of the same tile are glued) are to be glued along a common (reduced!) word w. By definition of the D_i's, the edges making up w do not originate from Γ_2.

As the edges of D_i originate from Γ_2, there is a lift $\varphi_i : D_i \to \Gamma_2$ (as noted above). Consider the two lifts $\varphi_i(w)$ and $\varphi_j(w)$. As the edges making up w do not originate from Γ_2, these two lifts are different. As w is reduced these lifts are immersions. So w is a piece. By assumption the length of w is at most $\Lambda < g/6$.

Now as D_i lifts to Γ_2, any boundary component of D_i goes to a closed path in Γ. This proves that the length of any boundary component of D_i is at least g.

So the tiles D_i satisfy the small cancellation condition with $\lambda = \Lambda/g < 1/6$. As they are tiles of a simply connected puzzle, by Corollary 14 they are simply connected.

Then by Lemma 13, the boundary of D is at least $1 - 6\lambda$ times the sum of the boundary lengths of the D_i's (considered as tiles). Since D is minimal, each D_i is as well, and as D_i is simply connected, by Corollary 12 it satisfies the isoperimetric inequality $|\partial D_i| \geqslant \alpha\,|D_i|$. So

$$|\partial D| \geqslant (1 - 6\lambda) \sum |\partial D_i| \geqslant \alpha(1 - 6\lambda) \sum |D_i| = \alpha(1 - 6\lambda)\,|D|$$

which shows the isoperimetric inequality for D, hence hyperbolicity.

For asphericity and the cohomological dimension (hence torsion-freeness), suppose that $\mathcal{C}$ is standard (so that Γ_2 is contractible) and that there exists a van Kampen-reduced spherical puzzle D, which we can assume to be inclusion-minimal in the sense that it contains no spherical subpuzzle. Define the D_i's as above. Either some D_i is spherical, in which case $D = D_i$ by inclusion-minimality of D, or all D_i's have non-empty boundary words. The former is ruled out by the following lemma:

Lemma 15.
Suppose that the set of paths read along faces of Γ_2 is standard. Let D be a non-empty spherical puzzle all edges of which originate from Γ_2. Then D is not van-Kampen reduced.

Proof of the lemma.
Let T be a maximal tree of Γ witnessing for standardness of the family of cycles. Homotope T to a point. This turns Γ_2 into a bouquet of circles with a face in each circle. Similarly, homotope to a point any edge of D coming from a suppressed edge of Γ. This way we turn D into a spherical van Kampen diagram with respect to the presentation of the fundamental group of Γ_2 (i.e. the trivial group) by $\langle\, c_1, \ldots, c_n \mid c_1 = e, \ldots, c_n = e \,\rangle$. But there is no reduced spherical van Kampen diagram with respect to this presentation, as can immediately be checked. □

Since by definition each D_i lifts to Γ_2 and since D (hence each D_i) is van Kampen-reduced, the lemma implies that no D_i is spherical. Hence the D_i's have non-trivial boundary words. So D can be viewed as a spherical puzzle with the boundary words of the D_i's as tiles. But we saw above that the D_i's (viewed as tiles) satisfy the small cancellation condition. So by Lemma 13 there is no spherical van Kampen-reduced puzzle w.r.t. these tiles.

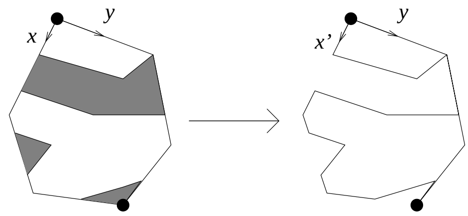
The computation of the Euler characteristic immediately follows, using that the cohomological dimension is at most 2.

The last assertions of the theorem follow easily from the assertions of Lemma 13. The smallest relation in the group presented by $\langle S \mid R \rangle$ is the boundary length of the smallest non-trivial puzzle, which by Lemma 13 is at least the smallest boundary length of the D_i's, which is at least the girth g. Similarly, any reduced word representing the trivial element in the group is read on the boundary of a van-Kampen reduced simply connected puzzle, thus contains as a subword at least one half of the boundary word of some D_i.

For the isometric embedding of Γ in the Cayley graph of the group, suppose that some geodesic path p in the graph (or in Γ_2) labelling a word x is equal to a shorter word y in the quotient. This means that there exists a puzzle D with boundary word xy^{-1}, made up of tiles with cycles of Γ as boundary words. We can take such a minimal couple (x, y) in the sense that the puzzle D has minimal number of tiles.

Now x is the word read on a path p_1 in the boundary word of D, which lifts to the geodesic path p in Γ_2 labelling x as well. Let f be a face of D which intersects p_1 along at least one edge. We say that f *originates from* Γ_2 *together with* x if the lift from f to Γ_2 coincides with the lift $p_1 \to p$ on the intersection of f with p_1.

Since we took (x, y) minimal (in the sense that D has minimal number of faces), we can assume that there is no face of D originating from Γ_2 together with x. Indeed, if there are some, they can be removed, yielding a new puzzle D' with fewer faces with boundary word $x'y^{-1}$. As x was geodesic in Γ_2, we have $|x'| \geqslant |x|$ so we still have $|y| < |x'|$, and thus the original couple (x, y) was not minimal. (Note that after this removal, x' and y may share a common initial or final subword, which can be truncated, as in the following picture in which black cells represent tiles originating from Γ_2 together with x.)



So we now assume that no tile of D originates from Γ_2 together with x. Now (if Γ contains no filaments) x is part of some cycle labelled by $w = xz$ of the graph. Since the path x is of minimal length, we have $|x| \leqslant |z|$. So $\left|xy^{-1}\right| < |w|$.

As Γ_2 is simply connected, there is a puzzle D' with boundary word w and which globally lifts to Γ_2 (all its edges originate from Γ_2). Define a new puzzle D'' by gluing D and D' along the word x. Now consider as above the partition $D'' = \cup D''_i$ with D''_i being maximal parts lifting to Γ_2. Since no tile of D adjacent to x originated from Γ_2 together with x, D' is exactly one of the D''_i.

So we have a puzzle bordering $z^{-1}y^{-1}$, which can be taken van Kampen-reduced since D and D' can be taken van Kampen-reduced and since there is no cancellation between D and D' (otherwise there would be a tile of D originating from Γ_2 together with x). But by Lemma 13, the boundary length $|z|+|y|$ of D'' is at least the boundary length of any D''_i, and in particular it is at least the boundary length of D', which is $|z| + |x|$. This is a contradiction since $|y| < |x|$.

This proves the theorem. $\square$

4 Further remarks

Remark 16.
The proof above gives an explicit isoperimetric constant when the set of relators taken is the set of all words read on cycles of the graph of length at most three times the diameter: in this case, any minimal simply connected puzzle satisfies the isoperimetric inequality

$$|\partial D| \geqslant g(1 - 6\Lambda/g)\,|D|\,/3$$

This explicit isoperimetric constant growing linearly with g (i.e. "homogeneous") can be very useful if one wants to apply such theorems as the local-global hyperbolic principle, which requires the isoperimetric constant to grow linearly with the sizes of the relators.

Remark 17.
The assumption that Γ is reduced can be relaxed a little bit, provided that some quasi-geodesicity assumption is granted, and that the definition of a piece is emended.

Redefine a *piece* to be a couple of words (w_1, w_2) such that both immerse in Γ and such that $w_1 = w_2$ in the free group. The *length* of a piece (w_1, w_2) is the maximal length of w_1 and w_2.

There are trivial pieces, for example if $w_1 = w_2$ and both have the same immersion. However, forbidding this is not enough: for example, if a word of the form $aa^{-1}w$ immerses in the graph, then $(aa^{-1}w, w)$ will be a piece.

A *trivial piece* is a piece (w_1, w_2) such that there exists a path p in Γ joining the beginning of the immersion of w_1 to the beginning of the immersion of w_2 such that p is labelled with a word equal to e in the free group.

The new theorem is as follows.

Theorem 18 (M. Gromov).
Let Γ be a finite non-filamenteous labelled graph. Let R be the set of words read on all cycles of Γ (or on a generating family of cycles). Let g be the girth of Γ and Λ be the length of the longest non-trivial piece of Γ.

Suppose that $\lambda = \Lambda/g$ is less than $1/6$.

Suppose that there exist a constant $A > 0$ such that any word w immersed in Γ of length at least L satisfies $\|w\| \geqslant A(|w| - L)$ for some $L < (1 - 6\lambda)g/2$.

Then the presentation $\langle\, S \mid R\,\rangle$ defines a hyperbolic, infinite, torsion-free group G, and (if R arises from a standard family of cycles) this presentation is aspherical (hence the cohomological dimension of G is at most 2). Moreover, the natural map of labelled graphs from Γ to the Cayley graph of G is a $(1/A, AL)$-quasi-isometry. The shortest *relation of G is of length at least $Ag/2$, and any reduced word equal to e in G contains as a subword the reduction of at least one half of a word read on a cycle of Γ.*

(Here $\|w\|$ is the length of the reduction of w; besides, in accordance with [GH90], by a (λ, c)-quasi-isometry we wean a map f such that $d(x,y)/\lambda - c \leqslant d(f(x), f(y)) \leqslant \lambda d(x,y) + c$.)

Remark 19.
The same kind of theorem holds if we use the $C(7)$ condition instead of the $C'(1/6)$ condition, but in this case there is no control on the radius of injectivity (shortest relation length).

Remark 20.
Using the techniques in [Del96] or [Oll04], the same kind of theorem should hold starting with any torsion-free hyperbolic group instead of the free group, provided that the girth of the graph is large enough w.r.t. the hyperbolicity constant, and that the labelling is quasi-geodesic. See [Oll].

Remark 21.
Theorem 1 can be extended when the graph is infinite, in which case we get a direct limit of torsion-free, dimension-2 hyperbolic groups (but generally not hyperbolic), in which the conclusions of small cancellation theory still hold but with the isoperimetric constant for van Kampen diagrams tending to 0. In this case the small cancellation assumption reads: any piece has length less than 1/6 times the minimal length of a cycle on which it appears.

References

[Bro82] K.S. Brown, *Cohomology of groups*, Graduate texts in Mathematics **87**, Springer (1982).

[CCH81] I.M. Chiswell, D.J. Collins, J. Huebschmann, *Aspherical group presentations*, Math. Z. **178** (1981), n° 1, 1–36.

[CH82] D.J. Collins, J. Huebschmann, *Spherical diagrams and identities among relations*, Math. Ann. **261** (1982), n° 2, 155–183.

[Del96] T. Delzant, *Sous-groupes distingués et quotients des groupes hyperboliques*, Duke Math. J. **83** (1996), n° 3, 661–682.

[Ger87] S.M. Gersten, *Reducible diagrams and equations over groups*, in *Essays in group theory*, ed. S.M. Gersten, Springer (1987), 15–73.

[Ghy03] É. Ghys, *Groupes aléatoires*, séminaire Bourbaki **916** (2003).

[GH90] É. Ghys, P. de la Harpe, *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math. **83**, Birkhäuser (1990).

[Gro87] M. Gromov, *Hyperbolic Groups*, in *Essays in group theory*, ed. S.M. Gersten, Springer (1987), 75–265.

[Gro01] M. Gromov, *Mesoscopic curvature and hyperbolicity*, in *Global differential geometry: the mathematical legacy of Alfred Gray*, ed. M. Fernández, J.A. Wolf, Contemporary Mathematics **288**, American Mathematical Society (2001), 58–69.

[Gro03] M. Gromov, *Random Walk in Random Groups*, GAFA, Geom. Funct. Anal. **13** (2003), n° 1, 73–146.

[HLS02] N. Higson, V. Lafforgue, G. Skandalis, *Counterexamples to the Baum-Connes conjecture*, Geom. Funct. Anal. **12** (2002), n° 2, 330–354.

[LS77] R.C. Lyndon, P.E. Schupp, *Combinatorial Group Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete **89**, Springer (1977).

[Oll04] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, GAFA, Geom. Funct. Anal. **14** (2004), n° 3, 595–679.

[Oll] Y. Ollivier, *Cayley graphs containing expanders, after Gromov*, expository manuscript (2003).

[OW] Y. Ollivier, D.T. Wise, *Kazhdan groups with infinite outer automorphism group*, to appear in Trans. Amer. Math. Soc.

[Sch73] P.E. Schupp, *A survey of small cancellation theory*, in W.W. Boone, F.B. Cannonito, R.C. Lyndon, eds, *Word problems*, Studies in logic and the foundations of mathematics **71**, North Holland Publ. Comp. (1973), 569–589.

[Sho91] H. Short et al., *Notes on word hyperbolic groups*, in *Group Theory from a Geometrical Viewpoint*, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991), 3–63.

[Sil03] L. Silberman, *Addendum to: "Random walk in random groups"*, GAFA, Geom. Funct. Anal. **13** (2003), n° 1, 147–177.

[Wis04] D.T. Wise, *Cubulating small cancellation groups*, GAFA, Geom. Funct. Anal. **14** (2004), n° 1, 150–214.

Kazhdan groups with infinite outer automorphism group

Ce texte est écrit en collaboration avec Dani Wise, de l'université McGill à Montréal. Après notre rencontre au Japon, Dani m'a invité à Montréal en août 2004, séjour au cours duquel nous avons rédigé ce travail. Ce texte est paru aux Transactions of the American mathematical society*, volume 359, n° 5 (2007), pp. 1959–1976.*

Kazhdan groups with infinite outer automorphism group

Yann Ollivier & Daniel T. Wise

Abstract

For each countable group Q we produce a short exact sequence $1 \to N \to G \to Q \to 1$ where G has a graphical $\frac{1}{6}$ presentation and N is f.g. and satisfies property T.

As a consequence we produce a group N with property T such that $\mathrm{Out}(N)$ is infinite.

Using the tools developed we are also able to produce examples of nonHopfian and non-coHopfian groups with property T.

One of our main tools is the use of random groups to achieve certain properties.

1 Introduction

1.1 Main statements

The object of this paper is to use a tool developed by Gromov to produce groups with property T that exhibit certain pathologies. Our main result is a property T variant of Rips' short exact sequence construction. We apply this to obtain a group with property T that has an infinite outer automorphism group. In two further examples, we produce non-Hopfian and non-coHopfian groups with property T.

In [Rip82], Rips gave an elementary construction which given a countable group Q produces a short exact sequence $1 \to N \to G \to Q \to 1$, where G is a $C'(\frac{1}{6})$ group and N is finitely generated. Rips used his construction to produce $C'(\frac{1}{6})$ presentations with various interesting properties, by lifting pathologies in Q to suitably reinterpreted pathologies in G.

In [Gro03], Gromov produced (random) groups with property T that have *graphical $\frac{1}{6}$ small cancellation* presentations. The graphical $\frac{1}{6}$ small cancellation condition is a generalization of the classical $C'(\frac{1}{6})$ condition (see [LS77]) to presentations determined by labelled graphs. We refer to Section 2 and [Oll-a] for a discussion of this property.

The mixture of these two tools yields the following in Section 3:

Theorem 1.1.
For each countable group Q, there is a short exact sequence $1 \to N \to G \to Q \to 1$ such that

1. *G is torsion-free,*
2. *G has a graphical $\frac{1}{6}$ presentation, and*
3. *N has property T.*
4. *G is finitely generated if Q is, and G is finitely presented if Q is.*

The graphical $\frac{1}{6}$ presentation retains enough properties of ordinary small cancellation theory to mix nicely with Rips' construction. However, we note that Theorem 1.1 cannot be obtained with G an ordinary $C'(\frac{1}{6})$ group, since finitely presented $C'(\frac{1}{6})$ groups act properly on a CAT(0) cube complex by [Wis04], and hence their infinite subgroups cannot have property T [NR97, NR98].

We apply Theorem 1.1 to obtain the following in Section 4:

Theorem 1.2.
Any countable group Q embeds in $\mathrm{Out}(N)$ for some group N with property T. In particular, there exists a group N with property T such that $\mathrm{Out}(N)$ is infinite.

The motivation is that, as proven by Paulin [Pau91], if H is word-hyperbolic and $|\mathrm{Out}(H)| = \infty$ then H splits over an infinite cyclic group, and hence H cannot have property T. The question of whether every group with property T has a finite outer automorphism group belongs to the list of open problems mentioned in de la Harpe and Valette's classical book on property T ([HV89], p. 134), was raised again by Alain Valette in his mathscinet review of [Pau91], and later appeared in a problem list from the 2002 meeting on property T at Oberwolfach.

It may be useful to remind the reader of the definition of Kazhdan's property T (see the excellent [HV89], or [Val04] for a more recent review). This is a property of linear unitary representations of a locally compact group G in the Hilbert space $\mathcal{H}$. Let ρ be a linear unitary representation of G in $\mathcal{H}$: It has an *invariant vector* if there exists a unitary $v \in \mathcal{H}$ such that, for all $g \in G$, we have $\rho(g)v = v$. It has *almost invariant vectors* if for any compact $K \subset G$, for every $\varepsilon > 0$, there exists a unitary vector $v \in \mathcal{H}$ such that, for all $g \in K$, $\|\rho(g)v - v\| \leqslant \varepsilon$ (so K "almost fixes" v). The group G has *property T* (or is a *Kazhdan group*) if any unitary representation of G having almost invariant vectors has an invariant vector.

Finally, we use the tools we developed to obtain the following two examples in Sections 6 and 5:

Theorem 1.3.
There exists a Kazhdan group G that is not Hopfian.

Theorem 1.4.
There exists a Kazhdan group G that is not coHopfian.

Various other attempts to augment Rips's construction have focused on strengthening the properties of G when Q is f.p. (e.g.: G is π_1 of a negatively curved complex [Wis98]; G is a residually finite $C'(\frac{1}{6})$ group [Wis03]; G is a subgroup of a right-angled Artin group, so $G \subset SL_n(\mathbb{Z})$ [HW]).

1.2 Random groups.

One key ingredient of our constructions is the use of random methods to provide examples of groups with particular properties. Random groups were introduced by Gromov, originally in [Gro87] and more and more extensively in [Gro93] and [Gro03]. The original motivation was and still is the study of "typical" properties of groups. But random groups now have applications. We refer to [Ghy04] and [Oll-c] for a general discussion of random groups.

We indeed use a result of [Gro03] (see also [Sil03]) providing a presentation of a group with property T satisyfing the graphical small cancellation property. In Section 7 we include a standalone proof of the results we need from [Gro03].

Żuk showed that random groups with "enough" relators very probably have property T [Zuk03]. However, these groups never satisfy any kind of $C'(\frac{1}{6})$ small cancellation property, since the number of relators is too large (typically, having enough relators to get property T creates pieces of relative length $> 1/3$ between the relators). So we rely on the more elaborate construction of [Gro03].

1.3 Acknowledgements

We would like to thank Yves de Cornulier and Frédéric Paulin for helpful comments on the text.

2 $Gr'(\frac{1}{6})$ graphs

2.1 Review of graphical α-condition $Gr'(\alpha)$

Throughout all this article, B is a bouquet of $m \geqslant 2$ circles whose edges are directed and labelled, so that m will be the number of generators of the group presentations we consider.

Let $\Gamma \looparrowright B$ be an immersed graph, and note that Γ has an induced labelling. That Γ immerses in B simply denotes the fact that the words carried by paths immersed in Γ are reduced.

By definition, the group G presented by $\langle B|\Gamma\rangle$, has generators the letters appearing on B, and relations consisting of all cycles appearing in F.

A *piece* P in Γ is an immersed path $P \looparrowright B$ which lifts to Γ in more than one way. We refer the reader to Figure 1 for a graph with some pieces in it. The a labels are indicated by white triangle, and the b labels are indicated by a black triangle. The bold path with label $aaba^{-1}b$ is a piece since it appears twice in the graph as indicated by the bold paths. Other pieces include the paths b^{10} and a^{11}.

Definition 2.1.
We say $\Gamma \looparrowright B$ satisfies the graphical α condition $Gr'(\alpha)$ *if for each piece P, and each cycle $C \to \Gamma$ such that $P \to \Gamma$ factors through $P \to C \to \Gamma$, we have $|P| < \alpha|C|$.*

The graphical α condition generalizes the usual $C'(\alpha)$: let F consist of the disjoint union of a set of cycles corresponding to the relators in a presentation. The

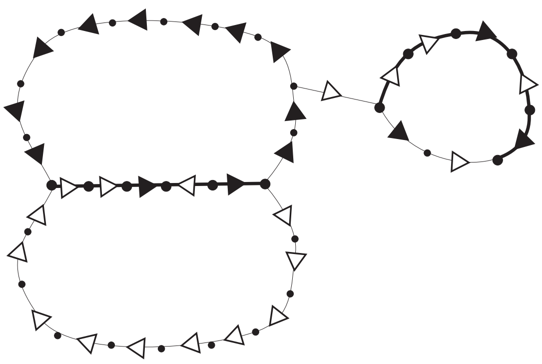


Figure 1:

graphical α condition is a case of a complicated but more general condition given by Gromov [Gro03].

The condition $Gr'(\frac{1}{6})$ implies that the group G is torsion-free, word-hyperbolic whenever the graph is finite, of dimension 2, just as the $C'(\frac{1}{6})$ condition [Oll-a]. The group is non-elementary except in some explicit degenerate cases (a hyperbolic group is called elementary if it is finite or virtually $\mathbb{Z}$).

There is also a slightly stronger version of this condition, in which we demand that the size of the pieces be bounded not by α times the size of any cycle containing the piece, but by α times the girth of Γ (recall the *girth* of a graph is the smallest length of a non-trivial closed path in it). We will sometimes directly prove this stronger version below, since it allows lighter notations.

A *disc van Kampen diagram* w.r.t. a graphical presentation is a van Kampen diagram every 2-cell of which is labelled by a closed path immersed in Γ. It is *reduced* if, first, it is reduced in the ordinary sense and if moreover, for any two adjacent 2-cells, the boundary word of their union does not embed as a closed path in Γ (otherwise, these two 2-cells can be replaced by a single one).

The following is easy to prove using the techniques in [Oll-a]:

Proposition 2.2.
Let Γ be a labelled graph satisfying $Gr'(\alpha)$ for $\alpha < \frac{1}{6}$. Let D be a disc van Kampen diagram with respect to the presentation defined by Γ. Suppose that D is reduced in the above sense. Then D locally satisfies the ordinary $C'(\alpha)$ small cancellation condition in the sense that two adjacent 2-cells of D share a common path of length at most α times the infimum of their lengths.

Proof.
Apply to D the decomposition of [Oll-a]: write D as a union of maximal parts the boundary word of which lifts to a cycle in Γ. Since D is reduced, this decomposition is trivial i.e. each single 2-face of D is such a maximal part (since otherwise we could replace all 2-cells in such a part by a single 2-cell with the same boundary word, which by definition lifts to a cycle in Γ). It is then proven in [Oll-a] that such maximal parts are in mutual $C'(\alpha)$ small cancellation with the adjacent parts. $\square$

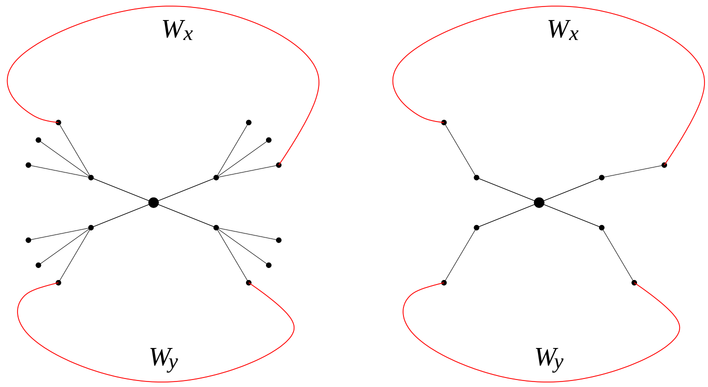


Figure 2:

Of course it is not true that the set of all boundary words of all 2-cells of a reduced diagram D satisfies the $C'(\alpha)$ condition: two non-adjacent 2-cells of D may have boundary words which partially intersect once lifted to Γ. But what we have is exactly what is needed to entail all the usual consequences of small cancellation like isoperimetry and the Greendlinger lemma.

2.2 Producing more $Gr'(\frac{1}{6})$ graphs

One useful feature of a presentation satisfying the ordinary $C'(\frac{1}{6})$ theory is that, provided that the relations are not "too dense" in a certain sense, more relations can be added to the presentation without violating the $C'(\frac{1}{6})$ condition.

In this subsection, we describe conditions on a $Gr'(\alpha)$ presentation such that additional relations can be added.

Proposition 2.3.
Let $\Gamma \looparrowright B$ satisfy the $Gr'(\alpha)$ condition and suppose there is an immersed path $W \to B$ such that $1 \leqslant |W| < \frac{\alpha}{2}\,\mathrm{girth}(\Gamma) - 1$, and W does not lift to Γ.

Then there is a set of closed immersed paths $C_i \looparrowright B : i \in \mathbb{N}$ such that the disjoint union $\Gamma' = \Gamma \sqcup_{i\in\mathbb{N}} C_i \looparrowright B$ satisfies the $Gr'(\alpha)$ condition.

Proof.
We first form an immersed labelled graph $A \looparrowright B$ as follows: Let D be the radius 2 ball at the basepoint of the universal cover $\tilde{B}$, and attach two copies W_x and W_y of the arc W along four distinct leaves of D as in Figure 2. (This can always be done avoiding the inverses of the initial and final letter of W, so that D immerses in B). Finally, we remove the finite trees that remain.

Observe that any path $P \looparrowright B$ that lifts to both A and Γ satisfies $|P| < \alpha\,\mathrm{girth}(\Gamma)$. Indeed, if P lifts to Γ, then P cannot contain W_x or W_y as a subpath, and hence $P = U_1U_2U_3$ where U_1 and U_3 are proper initial or terminal subpaths of a W-arc, and U_2 is a path in D, so $|P| \leqslant |U_1| + |U_2| + |U_3| \leqslant (|W|-1)+4+(|W|-1) = 2|W|+2 = 2(|W|+1) < 2\frac{\alpha}{2}\,\mathrm{girth}(\Gamma) = \alpha\,\mathrm{girth}(\Gamma)$.

Now let x and y be arbitrary labels. To any reduced word w in the letters $x^{\pm 1}$ and $y^{\pm 1}$ we can associate an immersed closed path $\varphi(w)$ in A by sending x to the based path in A containing W_x, and similarly for y.

Now for each $i \in \mathbb{N}$, let c_i denote the word $xy^{1000i+1}xy^{1000i+2}\cdots xy^{1000i+999}$. It is easily verified that for large enough values of 1000, the set of words $\langle x, y \mid c_i : i \in \mathbb{N}\rangle$ satisfies the $C'(\frac{\alpha}{2})$ condition.

Let $C_i \looparrowright B$ denote the corresponding closed immersed cycle $\varphi(c_i)$. Pieces in $\bigsqcup C_i$ are easily bounded in terms of pieces in $\langle x, y \mid c_i\,(i \in \mathbb{N})\rangle$, so that $\bigsqcup C_i$ satisfies the $Gr'(\alpha)$ (actually $C'(\alpha)$) condition.

Finally $\Gamma' = \Gamma \sqcup_{i\in\mathbb{N}} C_i$ satisfies the $Gr'(\alpha)$ condition since pieces that lift twice to Γ are bounded by assumption, and we have just bounded pieces that lift to Γ and to some C_i, and pieces that lift to some C_i and some C_j. $\square$

Remark 2.4.
The missing word condition in Γ ensures that the group presented by Γ is non-elementary. Indeed, the group presented by $\Gamma \sqcup_i C_i$ has infinite Euler characteristic (it is of dimension 2) and is thus non-elementary, so a fortiori the group presented by Γ is.

3 The T Rips construction

Let us now turn to the proof of the main theorem of this article. We use an intermediate construction due to Gromov.

Proposition 3.1.
There exists a finite graph Γ that immerses in a bouquet B of two circles such that:

1. *The group presented by $\langle B \mid \Gamma\rangle$ has property T.*
2. *$\Gamma \looparrowright B$ satisfies the $Gr'(\frac{1}{12})$ condition.*
3. *There is a path $W \looparrowright B$ with $1 \leqslant |W| < \frac{1}{24}\operatorname{girth}(\Gamma) - 1$ and W does not lift to Γ.*
4. *Γ has arbitrarily large girth.*

A proof of this is included in Section 7).

Theorem 1.1.
For each countable group Q, there is a short exact sequence $1 \to N \to G \to Q \to 1$ such that

1. *G is torsion-free,*
2. *G has a graphical $\frac{1}{6}$ presentation, and*
3. *N has property T and is non-trivial.*

4. Moreover, G is finitely generated if Q is, and finitely presented if Q is.

Proof.
Let Q be given by the following presentation:

$$\langle q_i : i \in I \mid R_j : j \in J \rangle$$

Let $\Gamma \looparrowright B$ be a graph provided by Proposition 3.1, where the edges of B are labelled by x and y. Let $\Gamma' = \Gamma \sqcup_n C_n$ be as in Proposition 2.3 with $\alpha = 1/12$.

The presentation for G will be the following:

$$\begin{aligned}\langle\ x, y,\ q_i\ (i \in I)\ \mid\ \Gamma,& \\ x^{q_i} = X_{i+},\ x^{q_i^{-1}} = X_{i-},\ y^{q_i} = Y_{i+},\ y^{q_i^{-1}} = Y_{i-}\ (i \in I),& \\ R_j = W_j\ (j \in J)\ \rangle& \quad (1)\end{aligned}$$

where superscripts denote conjugation, and where the X_{i+}, X_{i-}, Y_{i+}, Y_{i-}, and W_j are equal to paths corresponding to distinct C_n cycles of Γ', $|W_j| > 12|R_j|$ for each $j \in J$, and $|X_{i\pm}| > 36, |Y_{i\pm}| > 36$ for each $i \in I$.

The $\frac{1}{6}$ condition follows easily. Let us check, for example, that there is no $\frac{1}{6}$-piece between Γ and the relation $x^{q_i} = X_{i+}$. Since the q_i's do not appear as labels on Γ, any such $\frac{1}{6}$-piece would be either x or a subword of X_{i+}. The former is ruled out since $\mathrm{girth}(\Gamma) > 6$. The latter would provide a piece between Γ and X_{i+} (which is one of the C_n's); such a piece is by assumption of length at most $\frac{1}{12}|X_{i+}|$ which in turn is less than $\frac{1}{6}|x^{q_i} = X_{i+}|$ as needed. The other cases are treated similarly.

Now N is the subgroup of G generated by x and y. It is normal by construction of the presentation of G. Note that N has property T since it is a quotient of $\langle x, y \mid \Gamma\rangle$ which has property T by choice of Γ.

Finally, N is non-trivial: indeed, we can pick some cycle C_n which is a word in x, y and which will be in small cancellation with the rest of the presentation. This provides a word in x and y which is not trivial in the group. □

4 Kazhdan groups with infinite outer automorphism group

Theorem 1.2.
Any countable group Q embeds in $\mathrm{Out}(N)$ for some group N with property T.

In particular, there exists a group N with property T such that $\mathrm{Out}(N)$ is infinite.

Proof.
For $1 \to N \to G \to Q \to 1$, the group G acts by inner automorphisms on itself, so we have a homomorphism $G \to \mathrm{Aut}(N)$, and N obviously maps to $\mathrm{Inn}(N)$ so there is an induced homomorphism $Q = G/N \to \mathrm{Out}(N)$. Elements in the kernel of $Q \to \mathrm{Out}(N)$

are represented by elements $g \in G$ such that $m^g = m^n$ for some $n \in N$ and all $m \in N$. Thus gn^{-1} centralizes N.

First suppose that Q is finitely presented, so that G is as well.

In this case N is a non-elementary subgroup of the torsion-free word-hyperbolic group G, and hence N has a trivial centralizer. Indeed, N must contain a rank 2 free subgroup $\langle n_1, n_2 \rangle$ (see [GH90], p. 157). If a nontrivial element c centralizes N then $\langle c, n_1 \rangle$ and $\langle c, n_2 \rangle$ are both abelian, and hence infinite cyclic since G cannot contain a copy of $\mathbb{Z}^2$. Thus $n_1^{m_1} = c^{p_1}$ and $n_2^{m_2} = c^{p_2}$ for some $p_i, m_i \neq 0$. But then $n_1^{m_1}$ commutes with $n_2^{m_2}$ which is impossible.

Since the centralizer of N is trivial, we have $gn^{-1} = 1$, so $g \in N$, and hence $Q \to \mathrm{Out}(N)$ is injective.

The case when Q is not finitely presented reduces back to the previous one: Indeed, suppose that some element g of G lies in the centralizer of N. This is equivalent to stating that g commutes with x and y. But g can be written as a product of finitely many generators, and similarly the relations $[g, x] = 1$ and $[g, y] = 1$ are consequences of only finitely many relators, so that g still lies in the centralizer of N in a finite subpresentation of the presentation of G. □

Remark 4.1.
By adding some additional relations to N, the above argument was used in [BW05] to show that every countable group Q appears as $\mathrm{Out}(N)$ for some f.g. N, and that every f.p. Q appears as $\mathrm{Out}(N)$ where N is f.g. and residually finite (but property T did not appear there).

It appears likely that a more careful analysis along those lines, would show that every countable group arises as $\mathrm{Out}(N)$ where N has property T.

5 A Kazhdan group that is not coHopfian

Theorem 1.4.
There exists a Kazhdan group that is not coHopfian.

Proof.
Consider the group

$$G = \langle\, a, b, t \mid \Gamma,\ a^t = \varphi(a),\ b^t = \varphi(b) \,\rangle$$

where $\varphi(a)$ and $\varphi(b)$ are chosen so that $\Gamma \sqcup \varphi(a) \sqcup \varphi(b)$ satisfies $Gr'(\frac{1}{6})$ and $|\varphi(a)| > 3$, $|\varphi(b)| > 3$. (This is in fact a subpresentation of the presentation (1) used in the proof of Theorem 1.1.)

Clearly, the subgroup $K = \langle a, b \rangle$ is a Kazhdan group since it is a quotient of $\langle a, b \mid \Gamma \rangle$.

The map $K \to K$ induced by φ is clearly well-defined and injective since it arises from conjugation in the larger group G.

We will now show that φ is not surjective by verifying that $a \notin \langle \varphi(a), \varphi(b) \rangle$.

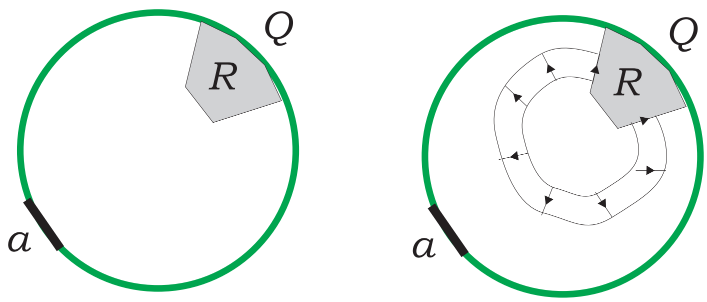


Figure 3:

We argue by contradiction: Suppose that a is equal in G to a word $W(\varphi(a), \varphi(b))$ in $\varphi(a)$ and $\varphi(b)$; we can choose W such that the disc diagram expressing this equality in the presentation for G has minimal area among all such choices. Note that since D is reduced and G is $Gr'(\frac{1}{6})$, D is a diagram satisfying the ordinary $C'(\frac{1}{6})$ condition in the sense of Proposition 2.2.

By Greendlinger's Lemma, (after ignoring trees possibly attached to ∂D) either D is a single 2-cell, or D has at least two 2-cells whose outer paths are the majority of their boundaries.

The first possibility is excluded by consideration of the presentation for G. In the second case, one such 2-cell R has outerpath Q not containing the special a-edge in ∂D, as illustrated on the left in Figure 3.

The boundary word of 2-cell R cannot be a word immersing in Γ. Indeed, since it has more than half its length on the boundary of D and this boundary bears a word in $\varphi(a)$ and $\varphi(b)$, this would contradict the small cancellation property of $\Gamma \sqcup \varphi(a) \sqcup \varphi(b)$. So R is a 2-cell expressing the equality $a^t = \varphi(a)$ or $b^t = \varphi(b)$. Moreover, since t does not appear on the boundary of D, the side of R on the boundary is the φ-side.

Since $t \notin \partial D$, we can find a t-annulus containing R as illustrated in the center of Figure 3. The t-edges in the figure are labelled by black triangles.

We now produce a new diagram D' with $\mathsf{Area}(D') < \mathsf{Area}(D)$. We do this by travelling around the t-annulus as on the right in Figure 3.

Observe that the small cancellation property implies that an edge in the $\varphi(a) \subset \partial R$ or $\varphi(b) \subset \partial R$ lines up with an edge in some $\varphi(a)$ or $\varphi(b)$ in ∂D, and at exactly the same position. So if the $\varphi(a)$ or $\varphi(b)$ of ∂R is not wholly contained in ∂D, after removing R the words on the paths from ∂D to the t-edges of R will cancel with corresponding subwords of $\varphi(a)$ and $\varphi(b)$ lying in the remaining part of ∂D.

This implies that, after removing the annulus, the boundary of D' is labelled (maybe after folding) by a word of the form $a = W'(\varphi(a), \varphi(b))$. But this is a contradiction since D was assumed to be minimal.

(Note that D' might touch the special a-edge, and D' might have some extra singular edges.) $\square$

6 A Kazhdan group that is not Hopfian

Definition 6.1.
Let B be a bouquet of circles, and let $\varphi : B \to B$. Let $A \to B$ be a map of graphs, then we let $\varphi(A) \to B$ be the new map of graphs where $\varphi(A)$ is obtained from A by substituting an arc $\varphi(e)$ for each edge e of A. That is, we replace the label on each edge of A by its image under φ.

Lemma 6.2.
Let Δ be a labelled graph satisfying $Gr'(\alpha)$ and $\alpha\,\mathrm{girth}(\Delta) \geqslant 1$. Suppose there is a path $P \looparrowright \Delta$ such that the edges in P all bear the same label a and such that P factors through a closed path $P \looparrowright C \looparrowright \Delta$. Then $|P| < 2\alpha\,|C|$.

Note that the assumption $\mathrm{girth}(\Delta) \geqslant 1/\alpha$ is not very strong: if $\alpha\,\mathrm{girth}(\Delta) \leqslant 1$ then a single letter can constitute a piece, which can result in various oddities. This lemma is false for trivial reasons if we remove this girth assumption: when $\mathrm{girth}(\Delta) = 1$ there are arbitrarily long homogeneous paths though $Gr'(0)$ may be satisfied.

Proof.
First, let us treat the trivial case when there is a length-1 loop bearing label a: this implies $\mathrm{girth}(\Delta) = 1$ so $\alpha = 1$ and the equality to show is trivial. The case $|P| = 1$ is trivial as well.

Second, suppose that there is no length-1 loop. Let P be a path labelled by a^s with $s \geqslant 2$. Then the two paths labelled by a^{s-1} obtained by removing the first and last edge of P respectively constitute a piece, and so we have $s - 1 < \alpha\,|C|$ so that $|P| = s < \alpha\,|C| + 1 \leqslant \alpha(|C| + \mathrm{girth}(\Delta)) \leqslant 2\alpha\,|C|$. □

Lemma 6.3.
Let Δ be a labelled graph satisfying $Gr'(\alpha)$ with $\alpha\,\mathrm{girth}(\Delta) \geqslant 1$, and let $\varphi : B \to B$ be induced by $a \mapsto a^n$ and $b \mapsto b^n$ for some $n \geqslant 1$. Then, for any $k \in \mathbb{N}$, $\varphi^k(\Delta)$ satisfies $Gr'(2\alpha)$.

(Once more the girth assumption discards some degenerate cases when a single edge can make a piece.)

Proof.
The reader should think of $\varphi^k(\Delta)$ as the n^k-subdivision of Δ where each a-edge is replaced by an arc of n^k a-edges and likewise for b-edges.

We begin by considering a homogeneous piece $P = a^r$ (or $P = b^r$ which is similar) occurring in some cycle $C \looparrowright \varphi^k(\Delta)$. Then P is a subpath of a path $\varphi^k(P')$ where $P' = a^{r'}$ is a path in Δ and P' occurs in a cycle C' corresponding to C.

By the previous lemma, $|P'| = r' < 2\alpha|C'|$ and so $|P| = r < 2\alpha n^k|C'| = 2\alpha|C|$.

We now consider the general case where P contains both a and b letters. We may assume that P is a maximal piece, in which case $P = W(a^{n^k}, b^{n^k})$ where $P' = W(a, b)$ is itself a corresponding piece in Δ. Everything scales by n^k i.e. $|P| = n^k|P'| < n^k\alpha|C'| = \alpha|C|$. □

Lemma 6.4.
Let Δ satisfy $Gr'(\alpha)$ and suppose that $\mathrm{girth}(\Delta) > 1/\alpha$. *Let n satisfy $n > s$ where s is the maximal length of a path a^s or b^s lifting to Δ. Let $\varphi : B \to B$ be induced by $a \mapsto a^n$ and $b \mapsto b^n$.*

Then $\bigsqcup_{k\geqslant 0} \varphi^k(\Delta)$ satisfies $Gr'(8\alpha)$.

Note that $s = \infty$ implies either $\mathrm{girth}(\Delta) = 1$ (which is excluded by assumption) or $\alpha = 1$ (by removing the first and last letter of an arbitrarily long a^s-path) in which case the affirmation is void. So we can suppose $s < \infty$.

Proof.
First, by the previous lemma, each $\varphi^k(\Delta)$ itself satisfies $Gr'(2\alpha)$.

We now consider a piece P between Δ and $\varphi^k(\Delta)$. Either $P \looparrowright \varphi^k(\Delta)$ is contained in two subdivided edges of $\varphi^k(\Delta)$ so $|P| < 2n^k$; or P contains an entire subdivided edge and hence an a^{n^k} (or b^{n^k}) subpath.

In the latter case when P contains an a^{n^k} or b^{n^k} subpath, since $P \looparrowright \Delta$ is a path in Δ then n^k is at most the maximal length of an a-path or b-path in Δ. But by hypothesis on n, this maximal length is bounded by n, and so $n^k < n$ which is impossible for $k \geqslant 1$.

In the former case, P is the product of at most two homogeneous paths (i.e. a-paths or b-paths) one of which has length $\geqslant \frac{1}{2}|P|$. Thus by Lemma 6.2, $\frac{1}{2}|P| < 2\alpha|C|$ for any cycle C in Δ containing P. So $|P| < 4\alpha|C|$ and so P cannot be a 4α-piece in Δ. Besides, suppose that P is included in a cycle C immersed in $\varphi^k(\Delta)$. Since $|P| < 2n^k$ and $|C| \geqslant \mathrm{girth}\,\varphi^k(\Delta) = n^k \,\mathrm{girth}(\Delta) \geqslant n^k/\alpha$ by assumption, P cannot consitute a 2α-piece in $\varphi^k(\Delta)$ either. (Note that the constant 4 is this reasoning is optimal: consider for Δ a circle of length 100 containing $aabb$ at one place, and some garbage for the rest; take $\alpha = (1+\varepsilon)/100$ so that the two a's do not form a piece. Then $\varphi(\Delta)$ contains some $aabb$ as well, so that this word constitutes a 4/100-piece in $\Delta \sqcup \varphi(\Delta)$.)

Finally, we consider pieces between $\varphi^k(\Delta)$ and $\varphi^{k'}(\Delta)$ where we can suppose $k' > k$. We have just proved that $\Delta \sqcup \varphi^{k'-k}(\Delta)$ satisfies $Gr'(4\alpha)$. We now apply Lemma 6.3 to see that $\varphi^k(\Delta) \sqcup \varphi^{k'}(\Delta) = \varphi^k\left(\Delta \sqcup \varphi^{k'-k}(\Delta)\right)$ satisfies $Gr'(8\alpha)$. ☐

Remark 6.5.
A generalization of Lemma 6.4 should hold with $\varphi(a)$ and $\varphi(b)$ appropriate small cancellation words instead of a^n and b^n.

Theorem 1.3.
There exists a Kazhdan group that is not Hopfian.

Proof.
Let G have the following presentation:

$$\langle\, a, b \mid \varphi^i(\Gamma),\ \varphi^i(a\varphi(C_1)),\ \varphi^i(b\varphi(C_2)),\ \varphi^i(\varphi(C_3))\ (i \geqslant 0)\,\rangle$$

where

1. $\Gamma \sqcup C_1 \sqcup C_2 \sqcup C_3$ satisfies the $Gr'(\alpha)$ condition with $\alpha = 1/2000$ (C_1, C_2 and C_3 arise from Proposition 2.3);
2. φ is defined by $\varphi(a) = a^n$ and $\varphi(b) = b^n$, for some n greater than the maximal length of an a-word or b-word in $\Gamma \sqcup C_1 \sqcup C_2 \sqcup C_3$;
3. $\mathrm{girth}(\Gamma \sqcup C_1 \sqcup C_2 \sqcup C_3) \geqslant 2000$.

Let $\Delta_0 = \bigsqcup_{k\geqslant 0} \varphi^k\, (\Gamma \sqcup C_1 \sqcup C_2 \sqcup C_3)$. By Lemma 6.4, this labelled graph satisfies $Gr'(8\alpha)$. As a subgraph of Δ_0, the graph $\Delta = \Gamma \sqcup \varphi(C_1) \sqcup \varphi(C_2) \sqcup C_3$ satisfies $Gr'(8\alpha)$ as well.

We now prove that $\Delta' = \Gamma \sqcup a\varphi(C_1) \sqcup b\varphi(C_2) \sqcup C_3$ is $Gr'(26\alpha)$. Let P be a piece involving the new a-edge or the new b-edge. Observe that $P = P_1 a P_2$ (or $P_1 b P_2$). Note that a new b (or new a) may lie in at most one of of P_1 or P_2. Thus P is the concatenation of at most 3 pieces in Δ together with the new a and possibly the new b. Consequently for any cycle C containing P we have $|P| < 24\alpha|C| + 2 \leqslant 26\alpha|C'|$ where we have used the hypothesis that $\alpha\,\mathrm{girth} \geqslant 1$.

We now apply Lemma 6.4 to see that $\Omega = \bigsqcup_{k\geqslant 0} \varphi^k(\Delta')$ satisfies $Gr'(208\alpha)$, and so does the presentation for G which is a subset of Ω.

Now φ obviously sends relations to relations and thus induces a well-defined map in G. This map is surjective since $a\varphi(C_1) =_G 1$ and $b\varphi(C_2) =_G 1$.

Finally φ is not injective since $\varphi(C_3) =_G 1$ but $C_3 \neq_G 1$. Indeed, C_3 is in small cancellation relative to the relators of G since both are included in Ω. $\square$

7 A T $Gr'(\frac{1}{6})$ graph with a missing word

A main point in this paper is the following, introduced by Gromov in [Gro03]:

Proposition 7.1.
For each $\alpha > 0$ and $\alpha' > 0$ there exists a finite graph Γ that immerses in a bouquet B of two circles such that:

1. *The group presented by $\langle\, B \mid \Gamma\, \rangle$ has property T.*
2. *$\Gamma \looparrowright B$ satisfies the $Gr'(\alpha)$ condition.*
3. *There is a path $W \looparrowright B$ with $1 \leqslant |W| \leqslant \alpha'\,\mathrm{girth}(\Gamma)$ and W does not lift to Γ.*

Moreover, the girth of Γ can be taken arbitrarily large.

This trivially implies Proposition 3.1. It also results from Remark 2.4 that the obtained group is non-trivial.

The goal of the introduction of such graphs in [Gro03] was to construct a group whose Cayley graph contains a family of expanders, in relation with the Baum-Connes conjecture (see also [Ghy04] and [Oll-b]). There, the construction is done starting not only with a free group but with an arbitrary hyperbolic group (compare [Oll04]), so that it can be iterated in order to embed a whole family of graphs.

Here we use this construction for purposes closer to combinatorial group theory. We do not need the full strength of the iterated construction; this section is devoted to the proof of the statements we need.

We will use the following fact, the credit of which can be shared between Lubotzky, Margulis, Phillips, Sarnack, Selberg. We refer to [Lub94] (Theorem 7.4.3 referring to Theorem 7.3.12), or to [DSV03].

Proposition 7.2.
For lots of $v \in \mathbb{N}$, there is a family of graphs $\Gamma_i : i \in \mathbb{N}$ such that the following hold:

1. *Each Γ_i is regular of valence v.*
2. $\inf_i \lambda_1(\Gamma_i) > 0$ *where λ_1 denotes the smallest non-zero eigenvalue of the discrete Laplacian Δ.*
3. $\operatorname{girth}(\Gamma_i) \longrightarrow \infty$.
4. $\exists C$ *such that* $\operatorname{diam}(\Gamma_i) \leqslant C \operatorname{girth}(\Gamma_i)$ *for all i.*

"Lots of v" means e.g. that this works at least for $v = p + 1$ with $p \geqslant 3$ prime ([Lub94], paragraph 1.2 refers to other constructions). This is irrelevant for our purpose.

We are going to use random labellings of subdivisions of the graphs Γ_i. Subdividing amounts to labelling each edge with a long word rather than just one letter, so that the small cancellation condition is more easily satisfied.

That the diameter of the graph is bounded by a constant times the girth reflects the fact that there are "not too many" relations added (compare the density model of random groups in [Gro93] or [Oll04]): this amounts to taking an arbitrarily small density.

To prove Proposition 7.1 we need two more propositions.

Proposition 7.3.
Given $v \in \mathbb{N}$, $\lambda_0 > 0$ and an integer $j \geqslant 1$ there exists an explicit g_0 such that if Γ is a graph with $\operatorname{girth}(\Gamma) \geqslant g_0$, $\lambda_1(\Gamma) \geqslant \lambda_0$ and every vertex of Γ has valency between 3 and v, then the random group defined through a random labelling of the j-subdivision Γ^j of Γ will have property T, with probability tending to 1 as the size of Γ tends to infinity.

This is proven in [Sil03] (Corollary 2.19 where d is our v, k is our number of generators m, and $|V|$ the size of the graph; in this reference, $\lambda(\Gamma)$ denotes the largest eigenvalue not equal to 1 of the averaging operator $1 - \Delta$, so that the inequalities between this λ and the first non-zero eigenvalue of Δ are reversed.)

In the next proposition and for the rest of this section, Γ^j denotes the j-subdivision of (the edges of) the graph Γ.

Proposition 7.4.
For any $v \in \mathbb{N}$, any $\alpha > 0$ and $\alpha' > 0$, for any $C \geqslant 1$, there exists an integer j_0 such that for any $j \geqslant j_0$, for any graph Γ satisfying the conditions:

1. *Each vertex of* Γ *is of valence at most* v*;*

2. *The girth of* Γ *is* g*;*

3. $\mathrm{diam}(\Gamma) \leqslant Cg$ *for all* i *(hence* Γ *is finite and connected);*

then the following properties hold with probability tending to 1 *as* $g \to \infty$*:*

1. *The folded graph* $\overline{\Gamma^j}$ *obtained by a random labelling of* Γ^j *satisfies the* $Gr'(\alpha)$ *condition.*

2. *There is a reduced word of length between* 1 *and* $\alpha' \operatorname{girth} \overline{\Gamma^j}$ *not appearing on any path in* $\overline{\Gamma^j}$*.*

This will be proven in the next sections (a sketch of proof can also be found in [Gro03]).

Let us now just gather propositions 7.2, 7.3 and 7.4.

Proof of Proposition 7.1.
Let α be the small cancellation constant to be achieved.

Apply Proposition 7.2 with some $v \in \mathbb{N}$ to get an infinite family of graphs Γ_i; let λ_0 be the lower bound on the spectral gap so obtained, and let C be as in this proposition. Let us denote by $\Gamma_{i(g)}$ the first graph in this family having girth at least g.

For the chosen $\alpha > 0$, let j and g be large enough for the conclusions of Proposition 7.4 to hold when applied to $\Gamma_{i(g)}$. Let g be still large enough (depending on j) so that the conclusions of Proposition 7.3 applied to this j hold. This provides a graph satisfying the three requirements of Proposition 7.1. □

7.1 Some simple properties of random words

Recall $m \geqslant 2$ is the number of generators we use. We denote by $\|w\|$ the norm in the free group of the word w, that is, the length of the associated reduced word.

Hereafter θ is the gross cogrowth of the free group (we refer to the paragraph "Growth, cogrowth, and gross cogrowth" in [Oll04] for basic properties). Basically, θ is the infimum of the real numbers so that the number of words of length ℓ which freely reduce to the trivial word is at most $(2m)^{\theta\ell}$ for all $\ell \in \mathbb{N}$. In particular, the probability that a random walk in the free group comes back at its origin at time ℓ is at most $(2m)^{-(1-\theta)\ell}$. Explicitly we have $(2m)^\theta = 2\sqrt{2m-1}$ [Kes59].

We state here some elementary properties having to deal with the behavior of reducing a random word. The first one is pretty intuitive.

Lemma 7.5.
Let W_ℓ *be a random word of length* ℓ *and let* $\overline{W}_\ell$ *be the associated reduced word. Then the law of* $\overline{W}_\ell$ *knowing its length* $\left|\overline{W}_\ell\right| = \|W_\ell\|$ *is the uniform law on all reduced words of this length.*

Proof of the lemma.
The group of automorphisms of the $2m$-regular tree preserving some basepoint acts transitively on the points at a given distance from the basepoint and preserves the law of the random walk beginning at this basepoint. □

The following is proven in [Oll04], Proposition 17.

Lemma 7.6.
Let W_ℓ be a random word of length ℓ. Then, for any $0 \leqslant L \leqslant \ell$ we have

$$\Pr(\|W_\ell\| \leqslant L) \leqslant (2m)^{-\ell(1-\theta)+\theta L}$$

Note that exponent vanishes for $L = \frac{1-\theta}{\theta}\ell < \ell$ (since $\theta > 1/2$). A slightly different, asymptotically stronger version of this lemma is the following.

Lemma 7.7.
Let W_ℓ be a random word of length ℓ. Then, for any L we have

$$\Pr(\|W_\ell\| \leqslant L) \leqslant \sqrt{\ell\,\frac{2m}{2m-1}}\,(2m)^{-(1-\theta)\ell}(2m-1)^{L/2}$$

Proof.
Let B_ℓ be the ball of radius ℓ centered at e in the free group. Let p_x^ℓ be the probability that $W_\ell = x$. We have

$$\begin{aligned} \mathbb{E}(2m-1)^{-\frac{1}{2}\|W_\ell\|} &= \sum_{x\in B_\ell} p_x^\ell\,(2m-1)^{-\frac{1}{2}\|x\|} \\ &\leqslant \sqrt{\sum_{x\in B_\ell}(p_x^\ell)^2}\,\sqrt{\sum_{x\in B_\ell}(2m-1)^{-\|x\|}} \end{aligned}$$

by the Cauchy-Schwarz inequality. But $\sum_{x\in B_\ell}(p_x^\ell)^2$ is exactly the probability of return to e at time 2ℓ of the random walk (condition by where it is at time ℓ) which is at most $(2m)^{-2(1-\theta)\ell}$. Besides, there are $(2m)(2m-1)^{k-1}$ elements of norm k in B_ℓ, so that $\sum_{x\in B_\ell}(2m-1)^{-\|x\|} = \sum_{0\leqslant k\leqslant \ell}(2m)(2m-1)^{k-1}(2m-1)^{-k} = \ell\frac{2m}{2m-1}$. So we get

$$\mathbb{E}(2m-1)^{-\frac{1}{2}\|W_\ell\|} \leqslant \sqrt{\ell\,\frac{2m}{2m-1}}\,(2m)^{-(1-\theta)\ell}$$

Now we simply apply the Markov inequality

$$\begin{aligned} \Pr(\|W_\ell\| \leqslant L) &= \Pr\left((2m-1)^{-\frac{1}{2}\|W_\ell\|} \geqslant (2m-1)^{-\frac{1}{2}L}\right) \\ &\leqslant (2m-1)^{\frac{1}{2}L}\,\mathbb{E}(2m-1)^{-\frac{1}{2}\|W_\ell\|} \end{aligned}$$

to get the conclusion. □

7.2 Folding the labelled graph

Labelling a graph by plain random words does generally not result in a reduced labelling. Nevertheless, we can always fold the resulting labelled graph. Here we show that in the circumstances needed for our applications, this folding is a quasi-isometry. This will allow a transfer of the Gr' small cancellation condition from the unfolded to the folded graph.

Proposition 7.8.
For any $\beta > 0$, for any $v \in \mathbb{N}$, for any $C \geqslant 1$, there exists an integer j_0 such that for any $j \geqslant j_0$, for any graph Γ satisfying the conditions:

1. *Vertices of Γ are of valency at most v.*

2. *$\operatorname{diam}(\Gamma) \leqslant Cg$ for all i, where g is the girth of Γ.*

then the folding map $\Gamma^j \to \overline{\Gamma^j}$ from a random labelling $\Gamma^j \to B$ to the associated reduced labelling $\overline{\Gamma^j} \looparrowright B$ is a $(\frac{\theta}{1-\theta}, \beta jg, gj)$ local quasi-isometry, with probability tending to 1 as $g \to \infty$.

We use the notation from [GH90] for local quasi-isometries: an (a, b, c) local quasi-isometry is a map f such that whenever $d(x, y) \leqslant c$ we have $\frac{1}{a}d(x, y) - b \leqslant d(f(x), f(y)) \leqslant ad(x, y) + b$. Here folding obviously decreases distances so that only the left inequality has to be checked.

Remark 7.9.
Below we will make repeated use of the following: The number of paths of length ℓ in Γ^j is at most $j^2\, v^{Cg+\ell/j}$. Indeed, the number of points in Γ is at most v^{Cg}, and once a point is chosen the number of paths of length k originating at it is at most v^k. Now specifying a path in the subdivision Γ^j amounts to specifying a path in Γ and giving two integers between 1 and j to specify the exact endpoints.

Proof.
Unwinding the definition of local quasi-isometries, we have to prove that any immersed path of length $\beta gj + \ell \leqslant gj$ in Γ^j is mapped onto a path of length at least $\frac{1-\theta}{\theta}\ell$ in $\overline{\Gamma^j}$.

By Remark 7.9, there are at most $j^2\, v^{Cg+g}$ paths of length gj in the subdivision Γ^j of Γ. Fix such a path, of length say $\beta gj + \ell$.

Since the length of the immersed path is at most $gj = \operatorname{girth}(\Gamma^j)$, the path does not travel twice along the same edge. Consequently, the labels appearing on this path are all chosen independently. Then by Lemma 7.6, the probability that its length after folding is less than $\frac{1-\theta}{\theta}\ell$ is less than

$$(2m)^{-(1-\theta)(\ell+\beta gj)+\theta\frac{1-\theta}{\theta}\ell} = (2m)^{-(1-\theta)\beta gj}$$

for this particular path. Since the number of choices for the path is at most $j^2\, v^{Cg+g}$, if j is large enough depending on C, β and θ, namely if $v^{C+1}(2m)^{-(1-\theta)\beta j} < 1$, then the probability that *there exists* a path violating our local quasi-isometry property will tend to 0 as $g \to \infty$. □

Corollary 7.10.
In the same circumstances, the girth of $\overline{\Gamma^j}$ is at least $\frac{1-\theta}{\theta} - \beta$ times that of Γ^j.

Proof.
Take a simple closed path p in $\overline{\Gamma_j}$. It is the image of a non-null-homotopic closed path q in Γ^j, whose length is by definition at least $gj = \operatorname{girth} \Gamma^j$. Let q' be the initial subpath of q of length gj. We can apply the local quasi-isometry statement to q', showing that its image p' has length at least $\frac{1-\theta}{\theta} gj - \beta gj$, which is thus a lower bound on the length of p. □

7.3 Pieces in the unfolded and folded graphs.

Here we show that under the circumstances above, the probability to get a long piece in the folded graph is very small.

Suppose again that we are given a graph Γ of degree at most v, of girth g and of diameter at most Cg. Consider its j-subdivision Γ^j endowed with a random labelling and let $\overline{\Gamma^j}$ be the associated folded labelled graph.

Let p, p' be two immersed paths in $\overline{\Gamma^j}$. Let q, q' be some preimages in Γ^i of p, p'. If p and p' are labelled by the same word, then q and q' will be labelled by some freely equal words, so that pieces come from pieces.

Note that in a graph labelled by non-reduced words, there are some "trivial pieces": e.g. if some aa^{-1} appears next to a word w, then $(w, aa^{-1}w)$ will be a piece. Such pieces disappear after folding the labelled graph; this is why we discard them in the following.

Proposition 7.11.
Let q, q' be two immersed paths in a graph Δ of girth g. Suppose that q and q' have length ℓ and ℓ' respectively, with ℓ and ℓ' at most $g/2$. Endow Δ with a random labelling. Suppose that after folding the graph, the paths q and q' are mapped to distinct paths. Then the probability that q and q' are labelled by two freely equal words is at most

$$C_{\ell,\ell'}(2m)^{-(1-\theta)(\ell+\ell')}$$

where $C_{\ell,\ell'}$ is a term growing subexponentially in $\ell + \ell'$.

Proof.
Let w and w' be the words labelling q and q' respectively.

First, assume that the images of q and q' in Δ are disjoint. Then the letters making up w and w' are chosen independently, and thus the word ww'^{-1} is a plain random word. Thus is this case the proposition is just a rewriting of the definition of θ.

Second, suppose that the paths do intersect in Δ: this results in lack of independence in the choice of the letters making up w and w' (the same problem is treated in a slightly different setting in [Oll04], section "Elimination of doublets"), which needs to be treated carefully. Since the length of these words is less than half the girth, the intersection in Δ is connected and we can write $w = u_1 u_2 u_3$, $w' = u'_1 u_2 u'_3$ where the u_i's are *independently chosen* random words (depending on relative orientation of w

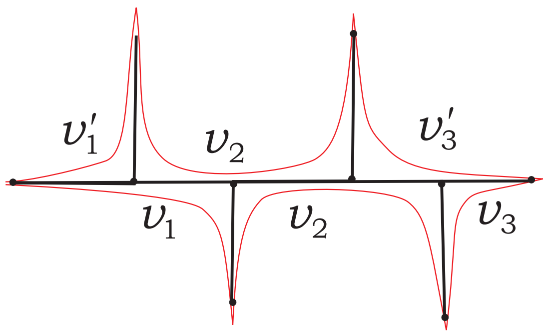


Figure 4:

and w', u_2^{-1} rather than u_2 may appear in w'). We can suppose that $u'_1u_1^{-1}$ is not freely trivial: otherwise the two paths start at the same point after folding, and so if $w = w'$ we also have $u'_3u_3^{-1} = e$ so that they also end at the same point after folding, but this is discarded by assumption. Likewise $u'_3u_3^{-1}$ is not freely trivial.

Let v_1, v_2, v_3, v'_1, v'_3 be the reduced words freely equal to u_1, u_2... respectively.

Lemma 7.5 tells us that the words v_1, v_2... are random reduced words. Now let us draw a picture expressing the equality $v_1v_2v_3 = v'_1v_2v'_3$:

Note that the two copies of v_2 have to be shifted relatively to each other, otherwise this means that $u'_1u_1^{-1}$ and $u'_3u_3^{-1}$ are freely trivial.

Let k be the length shared between the two copies of v_2. Now let us evaluate the probability of this situation knowing all the lengths of the words v_1, v_2, ... Conditionnally to their lengths, these words are uniformly chosen random reduced words by Lemma 7.5.

We begin with the two copies of v_2: though they are not chosen independently, since we know that they are shifted, adding letter after letter we see that the probability that they can glue along a subpath of length k is a most $1/(2m-1)^k$. Once v_2 is given, the words v_1, v_3, v'_1, v'_3 are all chosen independently of each other. The probability that they glue according to the picture is $1/(2m-1)^{L-k}$ where L is the total length of the picture. So the overall probability of such a gluing is $1/(2m-1)^L$.

We obviously have $\ell+\ell' = |v_1|+2|v_2|+|v_3|+|v'_1|+|v'_3| = 2L$. Now by Proposition 7.7 applied to all these words separately, the probability of achieving this value of $|v_1| + 2|v_2| + |v_3| + |v'_1| + |v'_3|$ is less than

$$C_{\ell,\ell'}\,(2m)^{-(1-\theta)(|u_1|+2|u_2|+|u_3|+|u'_1|+|u'_3|)}(2m-1)^{\frac{1}{2}\,2L} = C_{\ell,\ell'}\,(2m)^{-(1-\theta)(\ell+\ell')}(2m-1)^L$$

where $C_{\ell,\ell'}$ is a term growing subexponentially in $\ell+\ell'$.

So the overall probability of such a situation, taking into account the possibilities for L between 0 and $\ell+\ell'$, is at most

$$\sum_{0\leqslant L\leqslant \ell+\ell'} (2m-1)^{-L}\,C_{\ell+\ell'}\,(2m)^{-(1-\theta)(\ell+\ell')}(2m-1)^L$$

since we just proved above that $(2m-1)^{-L}$ is an upper bound for the probability of the situation knowing L. But this is equal to $C'_{\ell+\ell'}\,(2m)^{-(1-\theta)(\ell+\ell')}$ where $C'_{\ell+\ell'}$ is another term growing subexponentially in $\ell+\ell'$. $\square$

We are now ready to prove Proposition 7.4 stating that the $Gr'(\alpha)$ condition is satisfied with overwhelming probability. In order to avoid heavy notations, we will directly prove the stronger variant of the Gr' condition involving the girth instead of the length of cycles containing the pieces (see Section 2).

Proof of Proposition 7.4, small cancellation part.
Since ruling out small pieces rules out larger pieces as well, it is enough to work for small α.

Let $\bar{g}$ be the girth of $\overline{\Gamma^j}$. By Corollary 7.10, we can assume that $\bar{g} \geqslant \left(\frac{1-\theta}{\theta}-\beta\right) gj$ with overwhelming probability, for arbitrarily small β.

Let p, p' be two distinct immersed paths in $\overline{\Gamma^j}$ forming a α-piece; both p and p' are of length $\alpha\bar{g}$. Let q and q' be some immersed paths in Γ^j mapping to p and p'.

Suppose that the length of q (or q') is greater than $gj/2$. By applying the local quasi-isometry property to an initial subpath of q of length $gj/2$ we get that the length of p would be at least $\frac{1-\theta}{\theta}gj/2-\beta gj$. But the length of p is exactly $\alpha\bar{g}\leqslant\alpha gj$, so that if α and β are taken small enough (depending on θ) we get a contradiction. Hence, the length of q is at most $gj/2$, so that we are in a position to apply Proposition 7.11.

The length of q and q' is at least that of p and p' namely $\alpha\bar{g}$, and since $\bar{g}\geqslant\left(\frac{1-\theta}{\theta}-\beta\right) gj$, q and q' form a $\alpha\left(\frac{1-\theta}{\theta}-\beta\right)$-piece in Γ^j. Now Proposition 7.11 states that for fixed q and q' in Γ^j, the probability of this is at most $C_{gj}(2m)^{-(1-\theta)2gj\alpha\left(\frac{1-\theta}{\theta}-\beta\right)}$, where C_{gj} is a subexponential term in $|q|+|q'|\leqslant gj$.

By Remark 7.9, the number of choices for q and q' is at most $j^4\,v^{(2C+1)g}$. So the probability that one of these choices gives rise to a piece is at most

$$j^4\,v^{(2C+1)g}\,C_{gj}(2m)^{-(1-\theta)2gj\alpha\left(\frac{1-\theta}{\theta}-\beta\right)}$$

Now, if β is taken small enough (depending only on θ) and if j is taken large enough (depending on α,θ and C but not on g), namely if

$$v^{2C+1}\,(2m)^{-(1-\theta)2j\alpha\left(\frac{1-\theta}{\theta}-\beta\right)}<1$$

then this tends to 0 when g tends to infinity. $\square$

Proof of Proposition 7.4, missing word part.
We now prove that for any $\alpha'>0$, in the same circumstances, there exists a reduced word of length $\alpha'\,\mathrm{girth}(\overline{\Gamma^j})$ not appearing on any path in $\overline{\Gamma^j}$.

Let p be a simple path of length $\alpha'\bar{g}$ in $\overline{\Gamma^j}$. It is the image of some path q in Γ^j of length at least $\alpha'\bar{g}\geqslant\alpha'(\frac{1-\theta}{\theta}-\beta)gj$. But by Remark 7.9, the number of such paths in Γ^j is at most $j^2\,v^{Cg+\alpha'(\frac{1-\theta}{\theta}-\beta)g}$, whereas the total number of reduced words of this length is at least $(2m-1)^{\alpha'(\frac{1-\theta}{\theta}-\beta)gj}$. So if j is taken large enough (depending on α' and θ but not on g) that is if

$$v^{C+\alpha'(\frac{1-\theta}{\theta}-\beta)}<(2m-1)^{\alpha'(\frac{1-\theta}{\theta}-\beta)j}$$

then the possible reduced words outnumber the paths in $\overline{\Gamma^j}$ when $g \to \infty$, so that there has to be a missing word. □

8 Problems

Does there exist a finitely presented group N with property T such that $\mathrm{Out}(N)$ is infinite?

Let Q be a f.p. group with property T. Does there exist word-hyperbolic G with property T and f.g. normal subgroup N such that $Q = G/N$?

Do there exist f.p. Kazhdan groups which are not Hopfian or coHopfian?

Remark 8.1.
Yves de Cornulier informed us after learning of our work that he can construct explicit f.p. Kazhdan groups with infinite order outer automorphisms [Cor]. His examples use lattices in Lie groups.

References

[BW05] Inna Bumagin and Daniel T. Wise, *Every group is an outer automorphism group of a finitely generated group*, J. Pure Appl. Algebra **200** (2005), no. 1-2, 137–147.

[Cor] Yves de Cornulier, *Finitely presentable, non-Hopfian groups with Kazhdan's property (T) and infinite outer automorphism group*, To appear in Proc. Amer. Math. Soc.

[DSV03] Giuliana Davidoff, Peter Sarnak, and Alain Valette, *Elementary number theory, group theory, and Ramanujan graphs*, London Mathematical Society Student Texts, vol. 55, Cambridge University Press, Cambridge, 2003.

[HV89] Pierre de la Harpe and Alain Valette, *La propriété* (T) *de Kazhdan pour les groupes localement compacts*, Astérisque, no. 175, Soc. Math. Fr, 1989, With an appendix by M. Burger.

[GH90] É. Ghys and P. de la Harpe (eds.), *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Mathematics, vol. 83, Birkhäuser Boston Inc., Boston, MA, 1990, Papers from the Swiss Seminar on Hyperbolic Groups held in Bern, 1988.

[Ghy04] Étienne Ghys, *Groupes aléatoires (d'après Misha Gromov,...)*, Astérisque (2004), no. 294, 173–204, Séminaire Bourbaki, Vol. 2003/2004, Exp. 916.

[Gro87] M. Gromov, *Hyperbolic groups*, Essays in group theory, Math. Sci. Res. Inst. Publ., vol. 8, Springer, New York, 1987, pp. 75–263.

[Gro93] M. Gromov, *Asymptotic invariants of infinite groups*, Geometric group theory, Vol. 2 (Sussex, 1991), Cambridge Univ. Press, Cambridge, 1993, pp. 1–295.

[Gro03] M. Gromov, *Random walk in random groups*, Geom. Funct. Anal. **13** (2003), no. 1, 73–146.

[HW] Frédéric Haglund and Daniel T. Wise, *Special cube complexes*, Preprint, 2004.

[Kes59] Harry Kesten, *Symmetric random walks on groups*, Trans. Amer. Math. Soc. **92** (1959), 336–354.

[Lub94] Alexander Lubotzky, *Discrete groups, expanding graphs and invariant measures*, Progress in Mathematics, vol. 125, Birkhäuser Verlag, Basel, 1994, With an appendix by Jonathan D. Rogawski.

[LS77] Roger C. Lyndon and Paul E. Schupp, *Combinatorial group theory*, Springer-Verlag, Berlin, 1977, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 89.

[NR97] Graham Niblo and Lawrence Reeves, *Groups acting on* CAT(0) *cube complexes*, Geom. Topol. **1** (1997), approx. 7 pp. (electronic).

[NR98] Graham A. Niblo and Martin A. Roller, *Groups acting on cubes and Kazhdan's property (T)*, Proc. Amer. Math. Soc. **126** (1998), no. 3, 693–699.

[Oll04] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, GAFA, Geom. Funct. Anal. **14** (2004), no. 3, 595–679.

[Oll-a] Y. Ollivier, *On a small cancellation theorem of Gromov*, To appear in Bull. Belgian Math. Soc.

[Oll-b] Y. Ollivier, *Cayley graphs containing expanders, after Gromov*, Expository manuscript, 2003.

[Oll-c] Y. Ollivier, *A January 2005 invitation to random groups*, Expository manuscript, 2005.

[Pau91] Frédéric Paulin, *Outer automorphisms of hyperbolic groups and small actions on* **R**-*trees*, Arboreal group theory (Berkeley, CA, 1988), Math. Sci. Res. Inst. Publ., vol. 19, Springer, New York, 1991, pp. 331–343.

[Rip82] E. Rips, *Subgroups of small cancellation groups*, Bull. London Math. Soc. **14** (1982), no. 1, 45–47.

[Sil03] L. Silberman, *Addendum to "Random walk in random groups" by M. Gromov*, GAFA, Geom. Funct. Anal. **13** (2003), no. 1, 147–177.

[Val04] Alain Valette, *Nouvelles approches de la propriété (T) de Kazhdan*, Astérisque (2004), no. 294, 97–124, Séminaire Bourbaki, Vol. 2003/2004, Exp. 913.

[Wis98] Daniel T. Wise, *Incoherent negatively curved groups*, Proc. Amer. Math. Soc. **126** (1998), no. 4, 957–964.

[Wis03] Daniel T. Wise, *A residually finite version of Rips's construction*, Bull. London Math. Soc. **35** (2003), no. 1, 23–29.

[Wis04] Daniel T. Wise, *Cubulating small cancellation groups*, GAFA, Geom. Funct. Anal. **14** (2004), no. 1, 150–214.

[Zuk03] A. Żuk, *Property (T) and Kazhdan constants for discrete groups*, Geom. Funct. Anal. **13** (2003), no. 3, 643–670.

Ricci curvature of Markov chains on metric spaces

Les idées présentées dans ce texte ont émergé entre 2004 et 2007. Cet article est paru au volume 256, n°3 (2009) du Journal of Functional Analysis*, pp. 810–864. Les principaux résultats ont été annoncés au préalable dans la note* Ricci curvature of metric spaces *parue dans le volume 345, n°11 (2007) des* Comptes Rendus – Mathématique *de l'Académie des sciences de Paris, pp. 643–646.*

Ricci curvature of Markov chains on metric spaces

Yann Ollivier

Abstract

We define the coarse Ricci curvature of metric spaces in terms of how much small balls are closer (in Wasserstein transportation distance) than their centers are. This definition naturally extends to any Markov chain on a metric space. For a Riemannian manifold this gives back, after scaling, the value of Ricci curvature of a tangent vector. Examples of positively curved spaces for this definition include the discrete cube and discrete versions of the Ornstein–Uhlenbeck process. Moreover this generalization is consistent with the Bakry–Émery Ricci curvature for Brownian motion with a drift on a Riemannian manifold.

Positive Ricci curvature is shown to imply a spectral gap, a Lévy–Gromov–like Gaussian concentration theorem and a kind of modified logarithmic Sobolev inequality. The bounds obtained are sharp in a variety of examples.

Introduction

In Riemannian geometry, positively curved spaces in the sense of Ricci curvature enjoy numerous properties, some of them with very natural probabilistic interpretations. A basic result involving positive Ricci curvature is the Bonnet–Myers theorem bounding the diameter of the space via curvature; let us also mention Lichnerowicz's theorem for the spectral gap of the Laplacian (Theorem 181 in [Ber03]), hence a control on mixing properties of Brownian motion; and the Lévy–Gromov theorem for isoperimetric inequalities and concentration of measure [Gro86]. The scope of these theorems has been noticeably extended by Bakry–Émery theory [BE84, BE85], which highlights the analytic and probabilistic significance of Ricci curvature; in particular, they show that in positive Ricci curvature, a logarithmic Sobolev inequality holds. We refer to the nice survey [Lott] and the references therein for a discussion of the geometric interest of lower bounds on Ricci curvature and the need for a generalized notion of positive Ricci curvature for metric measure spaces.

Here we define a notion of Ricci curvature which makes sense for a metric space equipped with a Markov chain (or with a measure), and allows to extend the results above. Namely, we compare the transportation distance between the measures issuing from two given points to the distance between these points (Definition 3), so that Ricci curvature is positive if and only if the random walk operator is contracting on the space of probability measures equipped with this transportation distance

(Proposition 20). Thus, the techniques presented here are a metric version of the usual coupling method; namely, Ricci curvature appears as a refined version of Dobrushin's classical ergodic coefficient ([Dob56a, Dob56b], or e.g. Section 6.7.1 in [Bré99]) using the metric structure of the underlying space.

Our definition is very easy to implement on concrete examples. Especially, in ε-geodesic spaces, positive curvature is a local property (Proposition 19), as can be expected of a notion of curvature. As a result, we can test our notion in discrete spaces such as graphs. An example is the discrete cube $\{0,1\}^N$, which from the point of view of concentration of measure or convex geometry [MS86, Led01] behaves very much like the sphere S^N, and is thus expected to somehow have positive curvature.

Our notion enjoys the following properties: When applied to a Riemannian manifold equipped with (a discrete-time approximation of) Brownian motion, it gives back the usual value of the Ricci curvature of a tangent vector. It is consistent with the Bakry–Émery extension, and provides a visual explanation for the curvature contribution $-\nabla^{\mathrm{sym}} b$ of the drift term b in this theory. We are able to prove generalizations of the Bonnet–Myers theorem, of the Lichnerowicz spectral gap theorem and of the Lévy–Gromov isoperimetry theorem, as well as a kind of modified logarithmic Sobolev inequality. As a by-product, we get a new proof for Gaussian concentration and the logarithmic Sobolev inequality in the Lévy–Gromov or Bakry–Émery context (although with some loss in the numerical constants). We refer to Section 1.3 for an overview of the results.

Some of the results of this text have been announced in a short note [Oll07].

Historical remarks and related work. In the respective context of Riemannian manifolds or of discrete Markov chains, our techniques reduce, respectively, to Bakry–Émery theory or to a metric version of the coupling method. As far as I know, it had not been observed that these can actually be viewed as the same phenomenon.

From the discrete Markov chain point of view, the techniques presented here are just a version of the usual coupling method using the metric structure of the underlying space. Usually the coupling method involves total variation distance (see e.g. Section 6.7.1 in [Bré99]), which can be seen as a transportation distance with respect to the trivial metric. The coupling method is especially powerful in product or product-like spaces, such as spin systems. The work of Marton [Mar96a, Mar96b] emphasized the relationship between couplings and concentration of measure in product-like situations, so it is not surprising that we are able to get the same kind of results. The relationship between couplings and spectral gap is thoroughly explored in the works of Chen (e.g. [CL89, CW94, Che98]).

The contraction property of Markov chains in transportation distance seems to make its appearance in Dobrushin's paper [Dob70] (in which the current wide interest in transportation distances originates), and is implicit in the widely used "Dobrushin criterion" for spin systems [Dob70, DS85]. It later appears sporadically in the literature, as in Chen and Wang [CW94] (Thm. 1.9, as a tool for spectral gap estimates, using the coupling by reflection); at the very end of Dobrushin's notes [Dob96] (Dobrushin's study of the topic was stopped by his death); in Bubley and Dyer [BD97] for the particular case of product spaces, after Dobrushin; in the second edition of [Che04] (Section 5.3); in Djellout, Guillin and Wu [DGW04] in the context of dependent sequences of random variables to get Gaussian concentration results; in lecture notes by Peres [Per] and in [Sam] (p. 94). See also the related work mentioned below. However, the theorems exposed in our work are new.

From the Riemannian point of view, our approach boils down to contraction of the Lipschitz norm by the heat equation, which is one of the results of Bakry and Émery ([BE84, BE85], see also [ABCFGMRS00] and [RS05]). This latter property was suggested in [RS05] as a possible definition of a lower bound on Ricci curvature for diffusion operators in general spaces, though it does not provide an explicit value for Ricci curvature at a given point.

Another notion of lower bound on Ricci curvature, valid for length spaces equipped with a measure, has been simultaneously introduced by Sturm [Stu06], Lott and Villani [LV], and Ohta [Oht07] (see also [RS05] and [OV00]). It relies on ideas from optimal transportation theory and analysis of paths in the space of probability measures. Their definition keeps a lot of the properties traditionally associated with positive Ricci curvature, and is compatible with the Bakry–Émery extension. However, it has two main drawbacks. First, the definition is rather involved and difficult to check on concrete examples. Second, it is infinitesimal, and difficult to adapt to discrete settings [BS].

Related work. After having written a first version of this text, we learned that related ideas appear in some recent papers. Joulin [Jou07] uses contraction of the Lipschitz constant (under the name "Wasserstein curvature") to get a Poisson-type concentration result for continuous-time Markov chains on a countable space, at least in the bounded, one-dimensional case. Oliveira [Oli] considers Kac's random walk on $\mathrm{SO}(n)$; in our language, his result is that this random walk has positive coarse Ricci curvature, which allows him to improve mixing time estimates significantly.

Acknowledgements. I would like to thank Vincent Beffara, Fabrice Debbasch, Alessio Figalli, Pierre Pansu, Bruno Sévennec, Romain Tessera and Cédric Villani for numerous inspiring conversations about coarse geometry and Ricci curvature, as well as Thierry Bodineau, Djalil Chafaï, Aldéric Joulin, Shin-ichi Ohta and Roberto Imbuzeiro Oliveira for useful remarks on the manuscript and bibliographical references. Special thanks to Pierre Py for the two points x and y.

Notation. We use the symbol $\approx$ to denote equality up to a multiplicative universal constant (typically 2 or 4); the symbol $\sim$ denotes usual asymptotic equivalence. The word "distribution" is used as a synonym for "probability measure".

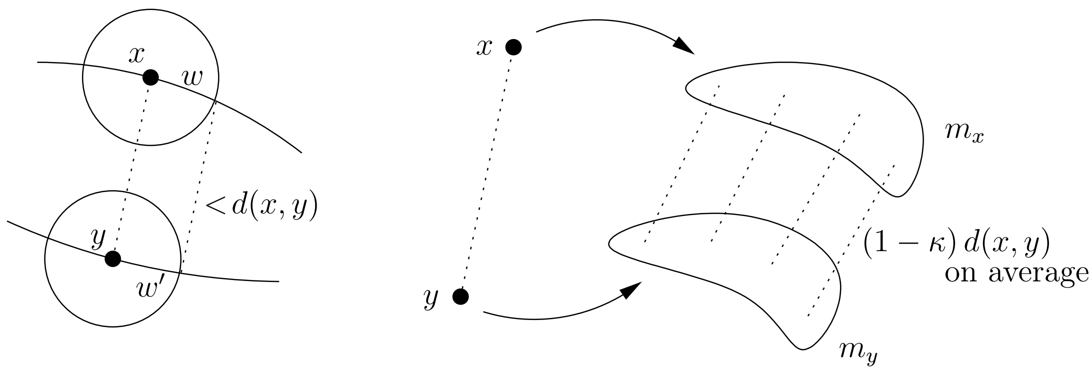
Here for simplicity we will mainly consider discrete-time processes. Similar definitions and results can be given for continuous time (see e.g. Section 3.3.4).

1 Definitions and statements

1.1 Coarse Ricci curvature

In Riemannian geometry, positive Ricci curvature is characterized [RS05] by the fact that "small spheres are closer (in transportation distance) than their centers are". More precisely, consider two very close points x, y in a Riemannian manifold, defining a tangent vector (xy). Let w be another tangent vector at x; let w' be the tangent vector at y obtained by parallel transport of w from x to y. Now if we follow the two geodesics issuing from x, w and y, w', in positive curvature the geodesics will get closer, and will part away in negative curvature. Ricci curvature along (xy) is this phenomenon, averaged on all directions w at x. If we think of a direction w at x as a point on a small sphere S_x centered at x, this shows that, on average, Ricci curvature controls whether the distance between a point of S_x and the corresponding point of S_y is smaller or larger than the distance $d(x, y)$.

In a more general context, we will use a probability measure m_x depending on x as an analogue for the sphere (or ball) S_x centered at x.



Definition 1.
Let (X, d) be a Polish metric space, equipped with its Borel σ-algebra.

A random walk m *on* X *is a family of probability measures* $m_x(\cdot)$ *on* X *for each* $x \in X$*, satisfying the following two technical assumptions:* (i) *the measure* m_x *depends measurably on the point* $x \in X$*;* (ii) *each measure* m_x *has finite first moment, i.e. for some (hence any)* $o \in X$*, for any* $x \in X$ *one has* $\int d(o, y)\, \mathrm{d}m_x(y) < \infty$.

Instead of "corresponding points" between two close spheres S_x and S_y, we will use transportation distances between measures. We refer to [Vil03] for an introduction to the topic. This distance is usually associated with the names of Kantorovich, Rubinstein, Wasserstein, Ornstein, Monge, and others (see [OEM] for a historical account); we stick to the simpler and more descriptive "transportation distance".

Definition 2.
Let (X, d) be a metric space and let ν_1, ν_2 be two probability measures on X. The L^1

transportation distance *between* ν_1 *and* ν_2 *is*

$$W_1(\nu_1, \nu_2) := \inf_{\xi \in \Pi(\nu_1, \nu_2)} \int_{(x,y) \in X \times X} d(x,y) \, \mathrm{d}\xi(x,y)$$

where $\Pi(\nu_1, \nu_2)$ *is the set of measures on* $X \times X$ *projecting to* ν_1 *and* ν_2.

Intuitively, $\mathrm{d}\xi(x,y)$ represents the mass that travels from x to y, hence the constraint on the projections of ξ, ensuring that the initial measure is ν_1 and the final measure is ν_2. The infimum is actually attained (Theorem 1.3 in [Vil03]), but the optimal coupling is generally not unique. In what follows, it is enough to choose one such coupling.

The data $(m_x)_{x \in X}$ allow to define a notion of curvature as follows: as in the Riemannian case, we will ask whether the measures m_x and m_y are closer or further apart than the points x and y are, in which case Ricci curvature will be, respectively, positive or negative.

Definition 3 (Coarse Ricci curvature).
Let (X, d) *be a metric space with a random walk* m*. Let* $x, y \in X$ *be two distinct points. The* coarse Ricci curvature of (X, d, m) along (xy) *is*

$$\kappa(x,y) := 1 - \frac{W_1(m_x, m_y)}{d(x,y)}$$

We will see below (Proposition 19) that in geodesic spaces, it is enough to know $\kappa(x,y)$ for close points x, y.

Geometers will think of m_x as a replacement for the notion of ball around x. Probabilists will rather think of this data as defining a Markov chain whose transition probability from x to y in n steps is

$$\mathrm{d}m_x^{*n}(y) := \int_{z \in X} \mathrm{d}m_x^{*(n-1)}(z) \, \mathrm{d}m_z(y)$$

where of course $m_x^{*1} := m_x$. Recall that a measure ν on X is *invariant* for this random walk if $\mathrm{d}\nu(x) = \int_y \mathrm{d}\nu(y) \mathrm{d}m_y(x)$. It is *reversible* if moreover, the detailed balance condition $\mathrm{d}\nu(x)\mathrm{d}m_x(y) = \mathrm{d}\nu(y)\mathrm{d}m_y(x)$ holds.

Other generalizations of Ricci curvature start with a metric measure space [Stu06, LV]. Here, as in Bakry–Émery theory, the measure appears as the invariant distribution of some process on the space (e.g. Brownian motion on a Riemannian manifold), which can be chosen in any convenient way. The following remark produces a random walk from a metric measure space, and allows to define the "Ricci curvature at scale ε" for any metric space.

Example 4 (ε-step random walk).
Let (X, d, μ) *be a metric measure space, and assume that balls in* X *have finite measure and that* $\operatorname{Supp}\mu = X$*. Choose some* $\varepsilon > 0$*. The* ε-step random walk *on* X*, starting*

at a point x, consists in randomly jumping in the ball of radius ε around x, with probability proportional to μ; namely, $m_x = \mu_{|B(x,\varepsilon)}/\mu(B(x,\varepsilon))$. (One can also use other functions of the distance, such as Gaussian kernels.)

As explained above, when (X, d) is a Riemannian manifold and m_x is the ε-step random walk with small ε, for close enough x, y this definition captures the Ricci curvature in the direction xy (up to some scaling factor depending on ε, see Example 7). In general there is no need for ε to be small: for example if X is a graph, $\varepsilon = 1$ is a natural choice.

If a continuous-time Markov kernel is given, one can also define a continuous-time version of coarse Ricci curvature by setting

$$\kappa(x,y) := -\frac{\mathrm{d}}{\mathrm{d}t}\frac{W_1(m_x^t, m_y^t)}{d(x,y)}$$

when this derivative exists (or take a lim inf), but for simplicity we will mainly work with the discrete-time version here. Indeed, for continuous-time Markov chains, existence of the process is already a non-trivial issue, even in the case of jump processes [Che04]. We will sometimes use our results on concrete continuous-time examples (e.g. $M/M/\infty$ queues in section 3.3.4), but only when they appear as an obvious limit of a discrete-time approximation.

One could use the L^p transportation distance instead of the L^1 one in the definition; however, this will make positive curvature a stronger assumption, and is never needed in our theorems.

Notation.
In analogy with the Riemannian case, when computing the transportation distance between measures m_x and m_y, we will think of $X \times X$ equipped with the coupling measure as a tangent space, and for $z \in X \times X$ we will write $x + z$ and $y + z$ for the two projections to X. So in this notation we have

$$\kappa(x,y) = \frac{1}{d(x,y)} \int (d(x,y) - d(x+z, y+z))\,\mathrm{d}z$$

where implicitly $\mathrm{d}z$ is the optimal coupling between m_x and m_y.

1.2 Examples

Example 5 ($\mathbb{Z}^N$ and $\mathbb{R}^N$).
Let m be the simple random walk on the graph of the grid $\mathbb{Z}^N$ equipped with its graph metric. Then for any two points $x, y \in \mathbb{Z}^d$, the coarse Ricci curvature along (xy) is 0.

Indeed, we can transport the measure m_x around x to the measure m_y by a translation of vector $y - x$ (and this is optimal), so that the distance between m_x and m_y is exactly that between x and y.

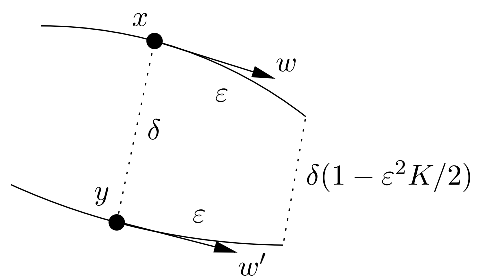
This example generalizes to the case of $\mathbb{Z}^N$ or $\mathbb{R}^N$ equipped with any distance and random walk which are translation-invariant (consistently with [LV]). For example, the triangular tiling of the plane has 0 curvature.

We now justify the terminology by showing that, in the case of the ε-step random walk on a Riemannian manifold, we get back the usual Ricci curvature (up to some scaling factor).

Proposition 6.
Let (X,d) be a smooth complete Riemannian manifold. Let v,w be unit tangent vectors at $x\in X$. Let $\varepsilon,\delta>0$. Let $y=\exp_x \delta v$ and let w' be the tangent vector at y obtained by parallel transport of w along the geodesic $\exp_x tv$. Then

$$d(\exp_x \varepsilon w, \exp_y \varepsilon w') = \delta\left(1-\frac{\varepsilon^2}{2}K(v,w)+O(\varepsilon^3+\varepsilon^2\delta)\right)$$

as $(\varepsilon,\delta)\to 0$. Here $K(v,w)$ is the sectional curvature in the tangent plane (v,w).



Example 7 (Riemannian manifold).
Let (X,d) be a smooth complete N-dimensional Riemannian manifold. For some $\varepsilon>0$, let the Markov chain m^ε be defined by

$$\mathrm{d}m_x^\varepsilon(y) := \frac{1}{\mathrm{vol}(B(x,\varepsilon))}\,\mathrm{d}\,\mathrm{vol}(y)$$

if $y\in B(x,\varepsilon)$, and 0 otherwise.

Let $x\in X$ and let v be a unit tangent vector at x. Let y be a point on the geodesic issuing from v, with $d(x,y)$ small enough. Then

$$\kappa(x,y)=\frac{\varepsilon^2\,\mathrm{Ric}(v,v)}{2(N+2)}+O(\varepsilon^3+\varepsilon^2 d(x,y))$$

The proof is postponed to Section 8; it is a refinement of Theorem 1.5 (condition (xii)) in [RS05], except that therein, the infimum of Ricci curvature is used instead of its value along a tangent vector. Basically, the value of $\kappa(x,y)$ is obtained by averaging Proposition 6 for w in the unit ball of the tangent space at x, which provides an upper bound for κ. The lower bound requires use of the dual characterization of transportation distance (Theorem 1.14 in [Vil03]).

Example 8 (Discrete cube).
Let $X = \{0,1\}^N$ be the discrete cube equipped with the Hamming metric (each edge is of length 1). Let m be the lazy random walk on the graph X, i.e. $m_x(x) = 1/2$ and $m_x(y) = 1/2N$ if y is a neighbor of x.

Let $x, y \in X$ be neighbors. Then $\kappa(x, y) = 1/N$.

This examples generalizes to arbitrary binomial distributions (see Section 3.3.3).

Here laziness is necessary to avoid parity problems: If no laziness is introduced, points at odd distance never meet under the random walk; in this case one would have to consider coarse Ricci curvature for points at even distance only.

Actually, since the discrete cube is a 1-geodesic space, one has $\kappa(x, y) \geqslant 1/N$ for any pair $x, y \in X$, not only neighbors (see Proposition 19).

Proof.
We can suppose that $x = 00\ldots0$ and $y = 10\ldots0$. For $z \in X$ and $1 \leqslant i \leqslant N$, let us denote by z^i the neighbor of z in which the i-th bit is switched. An optimal coupling between m_x and m_y is as follows: For $i \geqslant 2$, move x^i to y^i (both have mass $1/2N$ under m_x and m_y respectively). Now $m_x(x) = 1/2$ and $m_y(x) = 1/2N$, and likewise for y. So it is enough to move a mass $1/2 - 1/2N$ from x to y. All points are moved over a distance 1 by this coupling, except for a mass $1/2N$ which remains at x and a mass $1/2N$ which remains at y, and so the coarse Ricci curvature is at least $1/N$.

Optimality of this coupling is obtained as follows: Consider the function $f : X \to \{0,1\}$ which sends a point of X to its first bit. This is a 1-Lipschitz function, with $f(x) = 0$ and $f(y) = 1$. The expectations of f under m_x and m_y are $1/2N$ and $1 - 1/2N$ respectively, so that $1 - 1/N$ is a lower bound on $W_1(m_x, m_y)$.

A very short but less visual proof can be obtained with the L^1 tensorization property (Proposition 27). □

Example 9 (Ornstein–Uhlenbeck process).
Let $s \geqslant 0, \alpha > 0$ and consider the Ornstein–Uhlenbeck process in $\mathbb{R}^N$ given by the stochastic differential equation

$$\mathrm{d}X_t = -\alpha X_t \,\mathrm{d}t + s \,\mathrm{d}B_t$$

where B_t is a standard N-dimensional Brownian motion. The invariant distribution is Gaussian, of variance $s^2/2\alpha$.

Let $\delta t > 0$ and let the random walk m be the flow at time δt of the process. Explicitly, m_x is a Gaussian probability measure centered at $\mathrm{e}^{-\alpha\delta t}x$, of variance $s^2(1 - \mathrm{e}^{-2\alpha\delta t})/2\alpha \sim s^2\delta t$ for small δt.

Then the coarse Ricci curvature $\kappa(x, y)$ of this random walk is $1 - \mathrm{e}^{-\alpha\delta t}$, for any two $x, y \in \mathbb{R}^N$.

Proof.
The transportation distance between two Gaussian distributions with the same variance is the distance between their centers, so that $\kappa(x, y) = 1 - \frac{\left|\mathrm{e}^{-\alpha\delta t}x - \mathrm{e}^{-\alpha\delta t}y\right|}{|x-y|}$. □

Example 10 (Discrete Ornstein–Uhlenbeck).
Let $X = \{-N, -N+1, \dots, N-1, N\}$ and let m be the random walk on X given by

$$m_k(k) = 1/2, \qquad m_k(k+1) = 1/4 - k/4N, \qquad m_k(k-1) = 1/4 + k/4N$$

which is a lazy random walk with linear drift towards 0. The binomial distribution $\frac{1}{2^{2N}}\binom{2N}{N+k}$ is reversible for this random walk.

Then, for any two neighbors x, y in X, one has $\kappa(x,y) = 1/2N$.

Proof.
Exercise. □

Example 11 (Bakry–Émery).
Let X be an N-dimensional Riemannian manifold and F be a tangent vector field. Consider the differential operator

$$L := \frac{1}{2}\Delta + F.\nabla$$

associated with the stochastic differential equation

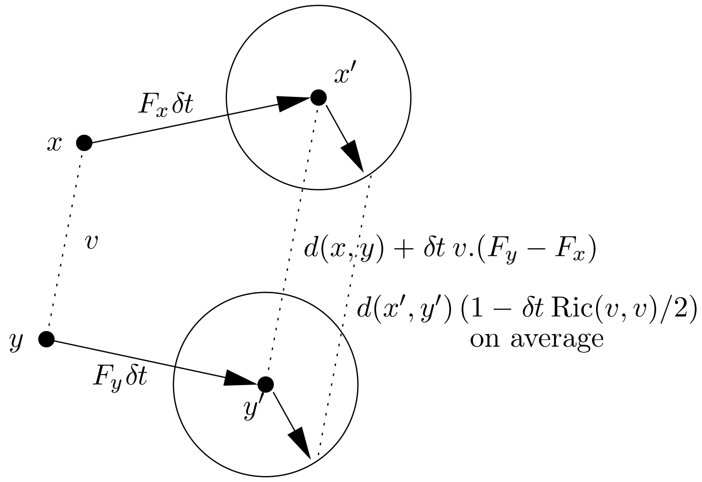
$$\mathrm{d}X_t = F\,\mathrm{d}t + \mathrm{d}B_t$$

where B_t is the Brownian motion in X. The Ricci curvature (in the Bakry–Émery sense) of this operator is $\frac{1}{2}\operatorname{Ric} - \nabla^{\mathrm{sym}} F$ where $\nabla^{\mathrm{sym}} F^{ij} := \frac{1}{2}(\nabla^i F^j + \nabla^j F^i)$ is the symmetrized of ∇F.

Consider the Euler approximation scheme at time δt for this stochastic equation, which consists in following the flow of F for a time δt and then randomly jumping in a ball of radius $\sqrt{(N+2)\delta t}$.

Let $x \in X$ and let v be a unit tangent vector at x. Let y be a point on the geodesic issuing from v, with $d(x,y)$ small enough. Then

$$\kappa(x,y) = \delta t \left(\frac{1}{2}\operatorname{Ric}(v,v) - \nabla^{\mathrm{sym}} F(v,v) + O(d(x,y)) + O(\sqrt{\delta t}) \right)$$



Proof.
First let us explain the normalization: Jumping in a ball of radius ε generates a variance $\varepsilon^2 \frac{1}{N+2}$ in a given direction. On the other hand, the N-dimensional Brownian motion has, by definition, a variance $\mathrm{d}t$ per unit of time $\mathrm{d}t$ in a given direction, so a proper discretization of Brownian motion at time δt requires jumping in a ball of radius $\varepsilon = \sqrt{(N+2)\delta t}$. Also, as noted in [BE85], the generator of Brownian motion is $\frac{1}{2}\Delta$ instead of Δ, hence the $\frac{1}{2}$ factor for the Ricci part.

Now the discrete-time process begins by following the flow F for some time δt. Starting at points x and y, using elementary Euclidean geometry, it is easy to see that after this, the distance between the endpoints behaves like $d(x,y)(1+\delta t\, v.\nabla_v F + O(\delta t^2))$. Note that $v.\nabla_v F = \nabla^{\mathrm{sym}} F(v,v)$.

Now, just as in Example 7, randomly jumping in a ball of radius ε results in a gain of $d(x,y)\frac{\varepsilon^2}{2(N+2)}\operatorname{Ric}(v,v)$ on transportation distances. Here $\varepsilon^2 = (N+2)\delta t$. So after the two steps of the process, the distance between the endpoints is

$$d(x,y)\left(1 - \frac{\delta t}{2}\operatorname{Ric}(v,v) + \delta t\, \nabla^{\mathrm{sym}} F(v,v)\right)$$

as needed, up to higher-order terms. □

Maybe the reason for the additional $-\nabla^{\mathrm{sym}} F$ in Ricci curvature à la Bakry–Émery is made clearer in this context: it is simply the quantity by which the flow of X modifies distances between two starting points.

It is clear on this example why reversibility is not fundamental in this theory: the antisymmetric part of the force F generates an infinitesimal isometric displacement. With our definition, combining the Markov chain with an isometry of the space has no effect whatsoever on curvature.

Example 12 (Multinomial distribution).
Consider the set $X = \{(x_0, x_1, \ldots, x_d),\ x_i \in \mathbb{N},\ \sum x_i = N\}$ viewed as the configuration set of N balls in $d+1$ boxes. Consider the process which consists in taking a ball at random among the N balls, removing it from its box, and putting it back at random in one of the $d+1$ boxes. More precisely, the transition probability from $(x_0, \ldots, x_d)$ to $(x_0, \ldots, x_i - 1, \ldots, x_j + 1, \ldots, x_d)$ (with maybe $i = j$) is $x_i/N(d+1)$. The multinomial distribution $\frac{N!}{(d+1)^N \prod x_i!}$ is reversible for this Markov chain.

Equip this configuration space with the metric $d((x_i),(x'_i)) := \frac{1}{2}\sum |x_i - x'_i|$ which is the graph distance w.r.t. the moves above. The coarse Ricci curvature of this Markov chain is $1/N$.

Proof.
Exercise (see also the discussion after Proposition 27). □

Example 13 (Geometric distribution).
Let the random walk on $\mathbb{N}$ be defined by the transition probabilities $p_{n,n+1} = 1/3$, $p_{n+1,n} = 2/3$ and $p_{0,0} = 2/3$. This random walk is reversible with respect to the geometric distribution $2^{-(n+1)}$. Then for $n \geqslant 1$ one has $\kappa(n, n+1) = 0$.

Proof.
The transition kernel is translation-invariant except at 0. □

Section 5 contains more material about this latter example and how non-negative coarse Ricci curvature sometimes implies exponential concentration.

Example 14 (Geometric distribution, 2).
Let the random walk on $\mathbb{N}$ be defined by the transition probabilities $p_{n,0} = \alpha$ and $p_{n,n+1} = 1 - \alpha$ for some $0 < \alpha < 1$. The geometric distribution $\alpha(1-\alpha)^n$ is invariant (but not reversible) for this random walk. The coarse Ricci curvature of this random walk is α.

Proof.
Exercise. □

Example 15 (δ-hyperbolic groups).
Let X be the Cayley graph of a non-elementary δ-hyperbolic group with respect to some finite generating set. Let k be a large enough integer (depending on the group) and consider the random walk on X which consists in performing k steps of the simple random walk. Let $x, y \in X$. Then $\kappa(x, y) = -2k/d(x, y)\,(1 + o(1))$ when $d(x, y)$ and k tend to infinity.

Note that $-2k/d(x, y)$ is the smallest possible value for $\kappa(x, y)$, knowing that the steps of the random walk are bounded by k.

Proof.
For z in the ball of radius k around x, and z' in the ball of radius k around y, elementary δ-hyperbolic geometry yields $d(z, z') = d(x, y) + d(x, z) + d(y, z') - (y, z)_x - (x, z')_y$ up to some multiple of δ, where $(\cdot, \cdot)$ denotes the Gromov product with respect to some basepoint [GH90]. Since this decomposes as the sum of a term depending on z only and a term depending on z' only, to compute the transportation distance it is enough to know the expectation of $(y, z)_x$ for z in the ball around x, and likewise for $(x, z')_y$. Using that balls have exponential growth, it is not difficult (see Proposition 21 in [Oll04]) to see that the expectation of $(y, z)_x$ is bounded by a constant, whatever k, hence the conclusion.

The same argument applies to trees or discrete δ-hyperbolic spaces with a uniform lower bound on the exponential growth rate of balls. □

Example 16 (Kac's random walk on orthogonal matrices, after [Oli]).
Consider the following random walk on the set of $N \times N$ orthogonal matrices: at each step, a pair of indices $1 \leqslant i < j \leqslant N$ is selected at random, an angle $\theta \in [0; 2\pi)$ is picked at random, and a rotation of angle θ is performed in the coordinate plane i, j. Equip $\mathrm{SO}(N)$ with the Riemannian metric induced by the Hilbert–Schmidt inner product $\mathrm{Tr}(a^*b)$ on its tangent space. It is proven in a preprint by Oliveira [Oli] that this random walk has coarse Ricci curvature $1 - \sqrt{1 - 2/N(N-1)} \sim 1/N^2$.

This is consistent with the fact that $\mathrm{SO}(N)$ has, as a Riemannian manifold, a positive Ricci curvature in the usual sense. However, from the computational point of view, Kac's random walk above is much nicer than either the Brownian motion or the ε-step random walk of Example 7. Oliveira uses his result to prove a new estimate $O(N^2 \ln N)$ for the mixing time of this random walk, neatly improving on previous estimates $O(N^4 \ln N)$ by Diaconis–Saloff-Coste and $O(N^{2.5} \ln N)$ by Pak–Sidenko; $\Omega(N^2)$ is an easy lower bound, see [Oli].

Example 17 (Glauber dynamics for the Ising model).
Let G be a finite graph. Consider the configuration space $X := \{-1, 1\}^G$ together with the energy function $U(S) := -\sum_{x\sim y\in G} S(x)S(y) - h\sum_x S(x)$ for $S \in X$, where $h \in \mathbb{R}$ is the external magnetic field. For some $\beta \geqslant 0$, equip X with the Gibbs distribution $\mu := \mathrm{e}^{-\beta U}/Z$ where as usual $Z := \sum_S \mathrm{e}^{-\beta U(S)}$. The distance between two states is defined as the number of vertices of G at which their values differ.

For $S \in X$ and $x \in G$, denote by S_{x+} and S_{x-} the states obtained from S by setting $S_{x+}(x) = +1$ and $S_{x-}(x) = -1$, respectively. Consider the following random walk on X (known as the Glauber dynamics*): at each step, a vertex $x \in G$ is chosen at random, and a new value for $S(x)$ is picked according to local equilibrium, i.e. $S(x)$ is set to 1 or -1 with probabilities proportional to $\mathrm{e}^{-\beta U(S_{x+})}$ and $\mathrm{e}^{-\beta U(S_{x-})}$ respectively (note that only the neighbors of x influence the ratio of these probabilities). The Gibbs distribution is reversible for this Markov chain.*

Then the coarse Ricci curvature of this Markov chain is at least

$$\frac{1}{|G|}\left(1 - v_{\max}\frac{\mathrm{e}^{\beta} - \mathrm{e}^{-\beta}}{\mathrm{e}^{\beta} + \mathrm{e}^{-\beta}}\right)$$

where $v_{\max}$ is the maximal valency of a vertex of G. In particular, if

$$\beta < \frac{1}{2}\ln\left(\frac{v_{\max}+1}{v_{\max}-1}\right)$$

then curvature is positive. Consequently, the critical β is at least this quantity.

This estimate for the critical temperature coincides with the one derived in [Gri67]. Actually, our argument generalizes to different settings (such as non-constant/negative values of the coupling J_{xy} between spins, or continuous spin spaces), and the positive curvature condition for the Glauber dynamics exactly amounts to the well-known one-site Dobrushin criterion [Dob70] (or to $G(\beta) < 1$ in the notation of [Gri67], Eq. (19)). By comparison, the exact value of the critical β for the Ising model on the regular infinite tree of valency v is $\frac{1}{2}\ln\left(\frac{v}{v-2}\right)$, which shows asymptotic optimality of this criterion. When block dynamics (see [Mar04]) are used instead of single-site updates, positive coarse Ricci curvature of the block dynamics Markov chain is equivalent to the Dobrushin–Shlosman criterion [DS85].

As shown in the rest of this paper, positive curvature implies several properties, especially, exponential convergence to equilibrium, concentration inequalities and a modified logarithmic Sobolev inequality. For the Glauber dynamics, the constants we get in these inequalities are essentially the same as in the infinite-temperature (independent) case, up to some factor depending on temperature which diverges when positive curvature ceases to hold. This is more or less equivalent to the main results of the literature under the Dobrushin–Shlosman criterion (see e.g. the review [Mar04]). Note however that in our setting we do not need the underlying graph to be $\mathbb{Z}^N$.

Proof.
Using Proposition 19, it is enough to bound coarse Ricci curvature for pairs of states at distance 1. Let S, S' be two states differing only at $x \in G$. We can suppose that $S(x) = -1$ and $S'(x) = 1$. Let m_S and $m_{S'}$ be the law of the step of the random walk issuing from S and S' respectively. We have to prove that the transportation distance between m_S and $m_{S'}$ is at most $1 - \frac{1}{|G|}\left(1 - v_{\max}\frac{e^{\beta}-e^{-\beta}}{e^{\beta}+e^{-\beta}}\right)$.

The measure m_S decomposes as $m_S = \frac{1}{|G|}\sum_{y\in G} m_S^y$, according to the vertex $y \in G$ which is modified by the random walk, and likewise for $m_{S'}$. To evaluate the transportation distance, we will compare m_S^y to $m_{S'}^y$.

If the step of the random walk consists in modifying the value of S at x (which occurs with probability $\frac{1}{|G|}$), then the resulting state has the same law for S and S', i.e. $m_S^x = m_{S'}^x$. Thus in this case the transportation distance is 0 and the contribution to coarse Ricci curvature is $1 \times \frac{1}{|G|}$.

If the step consists in modifying the value of S at some point y in G not adjacent to x, then the value at x does not influence local equilibrium at y, and so m_S^y and $m_{S'}^y$ are identical except at x. So in this case the distance is 1 and the contribution to coarse Ricci curvature is 0.

Now if the step consists in modifying the value of S at some point $y \in G$ adjacent to x (which occurs with probability $v_x/|G|$ where v_x is the valency of x), then the value at x does influence the law of the new value at y, by some amount which we now evaluate. The final distance between the two laws will be this amount plus 1 (1 accounts for the difference at x), and the contribution to coarse Ricci curvature will be negative.

Let us now evaluate this amount more precisely. Let $y \in G$ be adjacent to x. Set $a = e^{-\beta U(S_{y+})}/e^{-\beta U(S_{y-})}$. The step of the random walk consists in setting $S(y)$ to 1 with probability $\frac{a}{a+1}$, and to -1 with probability $\frac{1}{a+1}$. Setting likewise $a' = e^{-\beta U(S'_{y+})}/e^{-\beta U(S'_{y-})}$ for S', we are left to evaluate the distance between the distributions on $\{-1, 1\}$ given by $\left(\frac{a}{a+1}; \frac{1}{a+1}\right)$ and $\left(\frac{a'}{a'+1}; \frac{1}{a'+1}\right)$. It is immediate to check, using the definition of the energy U, that $a' = e^{4\beta}a$. Then, a simple computation shows that the distance between these two distributions is at most $\frac{e^{\beta}-e^{-\beta}}{e^{\beta}+e^{-\beta}}$. This value is actually achieved when y has odd valency, $h = 0$ and switching the value at x changes the majority of spin signs around y. (Our argument is suboptimal here when valency is even—a more precise estimation yields the absence of a phase transition on $\mathbb{Z}$.)

Combining these different cases yields the desired curvature evaluation. To convert this into an evaluation of the critical β, reason as follows: Magnetization, defined as $\frac{1}{|G|}\sum_{x\in G} S(x)$, is a $\frac{1}{|G|}$-Lipschitz function of the state. Now let μ_0 be the Gibbs measure without magnetic field, and μ_h the Gibbs measure with external magnetic field h. Use the Glauber dynamics with magnetic field h, but starting with an initial state picked under μ_0; Cor. 22 yields that the magnetization under μ_h is controlled by $\frac{1}{|G|}W_1(\mu_0, \mu_0 * m)/\kappa$ where κ is the coarse Ricci curvature, and $W_1(\mu_0, \mu_0 * m)$ is the transportation distance between the Gibbs measure μ_0 and the measure obtained from it after one step of the Glauber dynamics with magnetic field h; reasoning as above this transportation distance is easily bounded by $\frac{1}{|G|}\frac{\mathrm{e}^{\beta h}-\mathrm{e}^{-\beta h}}{\mathrm{e}^{\beta h}+\mathrm{e}^{-\beta h}}$, so that the derivative of magnetization w.r.t. h stays bounded when $|G| \to \infty$, which is the classical criterion used to define critical temperature. (Compare Eq. (22) in [Gri67].) $\square$

Further examples.
More examples can be found in Sections 3.3.3 (binomial and Poisson distributions), 3.3.4 ($M/M/\infty$ queues and generalizations), 3.3.5 (exponential tails), 3.3.6 (heavy tails) and 5 (geometric distributions on $\mathbb{N}$, exponential distributions on $\mathbb{R}^N$).

1.3 Overview of the results

Notation for random walks. Before we present the main results, we need to define some quantities related to the local behavior of the random walk: the *jump*, which will help control the diameter of the space, and the *coarse diffusion constant*, which controls concentration properties. Moreover, we define a notion of local dimension. The larger the dimension, the better for concentration of measure.

Definition 18 (Jump, diffusion constant, dimension).
Let the jump *of the random walk at x be*

$$J(x) := \mathbb{E}_{m_x} d(x, \cdot) = W_1(\delta_x, m_x)$$

Let the (coarse) diffusion constant *of the random walk at x be*

$$\sigma(x) := \left(\frac{1}{2}\iint d(y,z)^2 \,\mathrm{d}m_x(y)\,\mathrm{d}m_x(z)\right)^{1/2}$$

and, if ν is an invariant distribution, let

$$\sigma := \|\sigma(x)\|_{L^2(X,\nu)}$$

be the average diffusion constant.

Let also $\sigma_\infty(x) := \frac{1}{2}\operatorname{diam}\operatorname{Supp} m_x$ and $\sigma_\infty := \sup \sigma_\infty(x)$.

Let the local dimension *at x be*

$$n_x := \frac{\sigma(x)^2}{\sup\{\operatorname{Var}_{m_x} f, f : \operatorname{Supp} m_x \to \mathbb{R} \text{ 1-Lipschitz}\}}$$

and finally $n := \inf_x n_x$.

About this definition of dimension. Obviously $n_x \geqslant 1$. For the discrete-time Brownian motion on a N-dimensional Riemannian manifold, one has $n_x \approx N$ (see the end of Section 8). For the simple random walk on a graph, $n_x \approx 1$. This definition of dimension amounts to saying that in a space of dimension n, the typical variations of a (1-dimensional) Lipschitz function are $1/\sqrt{n}$ times the typical distance between two points. This is the case in the sphere S^n, in the Gaussian measure on $\mathbb{R}^n$, and in the discrete cube $\{0,1\}^n$. So generally one could define the "statistical dimension" of a metric measure space (X, d, μ) by this formula i.e.

$$\operatorname{StatDim}(X, d, \mu) := \frac{\frac{1}{2}\iint d(x,y)^2 \,\mathrm{d}\mu(x)\mathrm{d}\mu(y)}{\sup\{\operatorname{Var}_\mu f, f \text{ 1-Lipschitz}\}}$$

so that for each $x \in X$ the local dimension of X at x is $n_x = \operatorname{StatDim}(X, d, m_x)$. With this definition, $\mathbb{R}^N$ equipped with a Gaussian measure has statistical dimension N and local dimension $\approx N$, whereas the discrete cube $\{0,1\}^N$ has statistical dimension $\approx N$ and local dimension ≈ 1.

We now turn to the description of the main results of the paper.

Elementary properties. In Section 2 are gathered some straightforward results.

First, we prove (Proposition 19) that in an ε-geodesic space, a lower bound on $\kappa(x,y)$ for points x, y with $d(x,y) \leqslant \varepsilon$ implies the same lower bound for all pairs of points. This is simple yet very useful: indeed in the various graphs given above as examples, it was enough to compute the coarse Ricci curvature for neighbors.

Second, we prove equivalent characterizations of having coarse Ricci curvature uniformly bounded below: A space satisfies $\kappa(x,y) \geqslant \kappa$ if and only if the random walk operator is $(1-\kappa)$-contracting on the space of probability measures equipped with the transportation distance (Proposition 20), and if and only if the random walk operator acting on Lipschitz functions contracts the Lipschitz norm by $(1-\kappa)$ (Proposition 29). An immediate corollary is the existence of a unique invariant distribution when $\kappa > 0$.

The property of contraction of the Lipschitz norm easily implies, in the reversible case, that the spectral gap of the Laplacian operator associated with the random walk is at least κ (Proposition 30); this can be seen as a generalization of Lichnerowicz's theorem, and provides sharp estimates of the spectral gap in several examples. (A similar result appears in [CW94].)

In analogy with the Bonnet–Myers theorem, we prove that if coarse Ricci curvature is bounded below by $\kappa > 0$, then the diameter of the space is at most $2\sup_x J(x)/\kappa$ (Proposition 23). In case J is unbounded, we can evaluate instead the average distance to a given point x_0 under the invariant distribution ν (Proposition 24); namely, $\int d(x_0, y)\,\mathrm{d}\nu(y) \leqslant J(x_0)/\kappa$. In particular we have $\int d(x,y)\,\mathrm{d}\nu(x)\mathrm{d}\nu(y) \leqslant 2\inf J/\kappa$. These are L^1 versions of the Bonnet–Myers theorem rather than generalizations: from the case of manifolds one would expect $1/\sqrt{\kappa}$ instead of $1/\kappa$. Actually this L^1 version is sharp in all our examples except Riemannian manifolds; in Section 6 we investigate additional conditions for an L^2 version of the Bonnet–Myers theorem to hold.

Let us also mention some elementary operations preserving positive curvature: composition, superposition and L^1 tensorization (Propositions 25, 26 and 27).

Concentration results. Basically, if coarse Ricci curvature is bounded below by $\kappa > 0$, then the invariant distribution satisfies concentration results with variance $\sigma^2/n\kappa$ (up to some constant factor). This estimate is often sharp, as discussed in Section 3.3 where we revisit some of the examples.

However, the type of concentration (Gaussian, exponential, or $1/t^2$) depends on further local assumptions: indeed, the tail behavior of the invariant measure cannot be better than that of the local measures m_x. Without further assumptions, one only gets that the variance of a 1-Lipschitz function is at most $\sigma^2/n\kappa$, hence concentration like $\sigma^2/n\kappa t^2$ (Proposition 32). If we make the further assumption that the support of the measures m_x is uniformly bounded (i.e. $\sigma_\infty < \infty$), then we get mixed Gaussian-then-exponential concentration, with variance $\sigma^2/n\kappa$ (Theorem 33). The width of the Gaussian window depends on σ_∞, and on the rate of variation of the diffusion constant $\sigma(x)^2$.

For the case of Riemannian manifolds, simply considering smaller and smaller steps for the random walks makes the width of the Gaussian window tend to infinity, so that we recover full Gaussian concentration as in the Lévy–Gromov or Bakry–Émery context. However, for lots of discrete examples, the Gaussian-then-exponential behavior is genuine. Examples where tails are Poisson-like (binomial distribution, $M/M/\infty$ queues) or exponential are given in Sections 3.3.3 to 3.3.5. Examples of heavy tails (when $\sigma_\infty = \infty$) are given in 3.3.6.

We also get concentration results for the finite-time distributions m_x^{*k} (Remark 35).

Log-Sobolev inequality. Using a suitable non-local notion of norm of the gradient, we are able to adapt the proof by Bakry and Émery of a logarithmic Sobolev inequality for the invariant distribution. The gradient we use (Definition 41) is $(\mathrm{D}f)(x) := \sup_{y,z} \frac{|f(y)-f(z)|}{d(y,z)} \exp(-\lambda d(x,y) - \lambda d(y,z))$. This is a kind of "semi-local" Lipschitz constant for f. Typically the value of λ can be taken large at the "macroscopic" level; for Riemannian manifolds, taking smaller and smaller steps for the random walk allows to take $\lambda \to \infty$ so that we recover the usual gradient for smooth functions.

The inequality takes the form $\operatorname{Ent} f \leqslant C \int (\mathrm{D}f)^2/f \,\mathrm{d}\nu$ (Theorem 45). The main tool of the proof is the gradient contraction relation $\mathrm{D}(\mathrm{M}f) \leqslant (1-\kappa/2)\mathrm{M}(\mathrm{D}f)$ where M is the random walk operator (Theorem 44).

That the gradient is non-local, with a maximal possible value of λ, is consistent with the possible occurrence of non-Gaussian tails.

Exponential concentration and non-negative curvature. The simplest example of a Markov chain with zero coarse Ricci curvature is the simple random walk on $\mathbb{N}$ or $\mathbb{Z}$, for which there is no invariant distribution. However, we show that if furthermore there is a "locally attracting" point, then non-negative coarse Ricci curvature implies exponential concentration. Examples are the geometric distribution on $\mathbb{N}$ or the exponential distribution $e^{-|x|}$ on $\mathbb{R}^N$ associated with the stochastic differential equation $\mathrm{d}X_t = \mathrm{d}B_t - \frac{X_t}{|X_t|}\,\mathrm{d}t$. In both cases we recover correct orders of magnitude.

Gromov–Hausdorff topology. One advantage of our definition is that it involves only combinations of the distance function, and no derivatives, so that it is more or less impervious to deformations of the space. In Section 7 we show that coarse Ricci curvature is continuous for Gromov–Hausdorff convergence of metric spaces (suitably reinforced, of course, so that the random walk converges as well), so that having non-negative curvature is a closed property. We also suggest a loosened definition of coarse Ricci curvature, requiring that $W_1(m_x, m_y) \leqslant (1-\kappa)d(x,y) + \delta$ instead of $W_1(m_x, m_y) \leqslant (1-\kappa)d(x,y)$. With this definition, positive curvature becomes an *open* property, so that a space close to one with positive curvature has positive curvature.

2 Elementary properties

2.1 Geodesic spaces

The idea behind curvature is to use local properties to derive global ones. We give here a simple proposition expressing that in near-geodesic spaces such as graphs (with $\varepsilon = 1$) or manifolds (for any ε), it is enough to check positivity of coarse Ricci curvature for nearby points.

Proposition 19 (Geodesic spaces).
Suppose that (X, d) is ε-geodesic in the sense that for any two points $x, y \in X$, there exists an integer n and a sequence $x_0 = x, x_1, \ldots, x_n = y$ such that $d(x,y) = \sum d(x_i, x_{i+1})$ and $d(x_i, x_{i+1}) \leqslant \varepsilon$.

Then, if $\kappa(x,y) \geqslant \kappa$ for any pair of points with $d(x,y) \leqslant \varepsilon$, then $\kappa(x,y) \geqslant \kappa$ for any pair of points $x, y \in X$.

Proof.
Let (x_i) be as above. Using the triangle inequality for W_1, one has $W_1(m_x, m_y) \leqslant \sum W_1(m_{x_i}, m_{x_{i+1}}) \leqslant (1-\kappa)\sum d(x_i, x_{i+1}) = (1-\kappa)d(x,y)$. □

2.2 Contraction on the space of probability measures

Let $\mathcal{P}(X)$ by the space of all probability measures μ on X with finite first moment, i.e. for some (hence any) $o \in X$, $\int d(o,x)\,\mathrm{d}\mu(x) < \infty$. On $\mathcal{P}(X)$, the transportation distance W_1 is finite, so that it is actually a distance.

Let μ be a probability measure on X and define the measure

$$\mu * m := \int_{x\in X} \mathrm{d}\mu(x)\, m_x$$

which is the image of μ by the random walk. A priori, it may or may not belong to $\mathcal{P}(X)$.

The following proposition and its corollary can be seen as a particular case of Theorem 3 in [Dob70] (viewing a Markov chain as a Markov field on $\mathbb{N}$). Equivalent statements also appear in [Dob96] (Proposition 14.3), in the second edition of [Che04] (Theorem 5.22), in [DGW04] (in the proof of Proposition 2.10), in [Per] and in [Oli].

Proposition 20 (W_1 contraction).
Let (X, d, m) be a metric space with a random walk. Let $\kappa \in \mathbb{R}$. Then we have $\kappa(x, y) \geqslant \kappa$ for all $x, y \in X$, if and only if for any two probability distributions $\mu, \mu' \in \mathcal{P}(X)$ one has

$$W_1(\mu * m, \mu' * m) \leqslant (1 - \kappa) W_1(\mu, \mu')$$

Moreover in this case, if $\mu \in \mathcal{P}(X)$ then $\mu * m \in \mathcal{P}(X)$.

Proof.
First, suppose that convolution with m is contracting in W_1 distance. For some $x, y \in X$, let $\mu = \delta_x$ and $\mu' = \delta_y$ be the Dirac measures at x and y. Then by definition $\delta_x * m = m_x$ and likewise for y, so that $W_1(m_x, m_y) \leqslant (1-\kappa) W_1(\delta_x, \delta_y) = (1-\kappa) d(x, y)$ as required.

The converse is more difficult to write than to understand. For each pair (x, y) let ξ_{xy} be a coupling (i.e. a measure on $X \times X$) between m_x and m_y witnessing for $\kappa(x, y) \geqslant \kappa$. According to Corollary 5.22 in [Vil08], we can choose ξ_{xy} to depend measurably on the pair (x, y).

Let Ξ be a coupling between μ and μ' witnessing for $W_1(\mu, \mu')$. Then $\int_{X \times X} \mathrm{d}\Xi(x, y)\, \xi_{xy}$ is a coupling between $\mu * m$ and $\mu' * m$ and so

$$\begin{aligned}
W_1(\mu * m, \mu' * m) &\leqslant \int_{x,y} d(x, y)\, \mathrm{d}\left\{ \int_{x',y'} \mathrm{d}\Xi(x', y')\, \xi_{x'y'} \right\}(x, y) \\
&= \int_{x,y,x',y'} \mathrm{d}\Xi(x', y')\, \mathrm{d}\xi_{x'y'}(x, y)\, d(x, y) \\
&\leqslant \int_{x',y'} \mathrm{d}\Xi(x', y')\, d(x', y')(1 - \kappa(x', y')) \\
&\leqslant (1 - \kappa) W_1(\mu, \mu')
\end{aligned}$$

by the Fubini theorem applied to $d(x, y)\, \mathrm{d}\Xi(x', y')\, \mathrm{d}\xi_{x',y'}(x, y)$.

To see that in this situation $\mathcal{P}(X)$ is preserved by the random walk, fix some origin $o \in X$ and note that for any $\mu \in \mathcal{P}(X)$, the first moment of $\mu * m$ is $W_1(\delta_o, \mu * m) \leqslant W_1(\delta_o, m_o) + W_1(m_o, \mu * m) \leqslant W_1(\delta_o, m_o) + (1 - \kappa) W_1(\delta_o, \mu)$. Now $W_1(\delta_o, \mu) < \infty$ by assumption, and $W_1(\delta_o, m_o) < \infty$ by Definition 1. □

As an immediate consequence of this contracting property we get:

Corollary 21 (W_1 convergence).
Suppose that $\kappa(x, y) \geqslant \kappa > 0$ for any two distinct $x, y \in X$. Then the random walk has a unique invariant distribution $\nu \in \mathcal{P}(X)$.

Moreover, for any probability measure $\mu \in \mathcal{P}(X)$*, the sequence* $\mu * m^{*n}$ *tends exponentially fast to* ν *in* W_1 *distance. Namely*

$$W_1(\mu * m^{*n}, \nu) \leqslant (1-\kappa)^n \, W_1(\mu, \nu)$$

and in particular

$$W_1(m_x^{*n}, \nu) \leqslant (1-\kappa)^n J(x)/\kappa$$

The last assertion follows by taking $\mu = \delta_x$ and noting that $J(x) = W_1(\delta_x, m_x)$ so that $W_1(\delta_x, \nu) \leqslant W_1(\delta_x, m_x) + W_1(m_x, \nu) \leqslant J(x) + (1-\kappa) W_1(\delta_x, \nu)$, hence $W_1(\delta_x, \nu) \leqslant J(x)/\kappa$.

This is useful to provide bounds on mixing time. For example, suppose that X is a graph; since the total variation distance between two measures μ, μ' is the transportation distance with respect to the trivial metric instead of the graph metric, we obviously have $|\mu - \mu'|_{\mathrm{TV}} \leqslant W_1(\mu, \mu')$, hence the corollary above yields the estimate $\left|m_x^{*t} - \nu\right|_{\mathrm{TV}} \leqslant (\operatorname{diam} X)\,(1-\kappa)^t$ for any $x \in X$. Applied for example to the discrete cube $\{0,1\}^N$, with $\kappa = 1/N$ and diameter N, this gives the correct estimate $O(N \ln N)$ for mixing time in total variation distance, whereas the traditional estimate based on spectral gap and passage from L^2 to L^1 norm gives $O(N^2)$. Also note that the pointwise bound $\left|m_x^{*t} - \nu\right|_{\mathrm{TV}} \leqslant (1-\kappa)^t J(x)/\kappa$ depends on local data only and requires no knowledge of the invariant measure (compare [DS96]) or diameter; in particular it applies to infinite graphs.

Another immediate interesting corollary is the following, which allows to estimate the average of a Lipschitz function under the invariant measure, knowing some of its values. This is useful in concentration theorems, to get bounds not only on the deviations from the average, but on what the average actually is.

Corollary 22.
Suppose that $\kappa(x,y) \geqslant \kappa > 0$ *for any two distinct* $x, y \in X$*. Let* ν *be the invariant distribution.*

Let f *be a 1-Lipschitz function. Then, for any distribution* μ*, one has* $|\mathbb{E}_\nu f - \mathbb{E}_\mu f| \leqslant W_1(\mu, \mu * m)/\kappa$*.*

In particular, for any $x \in X$ *one has* $|f(x) - \mathbb{E}_\nu f| \leqslant J(x)/\kappa$*.*

Proof.
One has $W_1(\mu * m, \nu) \leqslant (1-\kappa) W_1(\mu, \nu)$. Since by the triangle inequality, $W_1(\mu * m, \nu) \geqslant W_1(\mu, \nu) - W_1(\mu, \mu * m)$, one gets $W_1(\mu, \nu) \leqslant W_1(\mu, \mu * m)/\kappa$. Now if f is a 1-Lipschitz function, for any two distributions μ, μ' one has $\left|\mathbb{E}_\mu f - \mathbb{E}_{\mu'} f\right| \leqslant W_1(\mu, \mu')$ hence the result.

The last assertion is simply the case when μ is the Dirac measure at x. $\square$

2.3 L^1 Bonnet–Myers theorems

We now give a weak analogue of the Bonnet–Myers theorem. This result shows in particular that positivity of coarse Ricci curvature is a much stronger property than a spectral gap bound: there is no coarse Ricci curvature analogue of a family of expanders.

Proposition 23 (L^1 Bonnet–Myers).
Suppose that $\kappa(x,y) \geqslant \kappa > 0$ for all $x, y \in X$. Then for any $x, y \in X$ one has

$$d(x,y) \leqslant \frac{J(x) + J(y)}{\kappa(x,y)}$$

and in particular

$$\operatorname{diam} X \leqslant \frac{2 \sup_x J(x)}{\kappa}$$

Proof.
We have $d(x,y) = W_1(\delta_x, \delta_y) \leqslant W_1(\delta_x, m_x) + W_1(m_x, m_y) + W_1(m_y, \delta_y) \leqslant J(x) + (1 - \kappa)d(x,y) + J(y)$ hence the result. □

This estimate is not sharp at all for Brownian motion in Riemannian manifolds (since $J \approx \varepsilon$ and $\kappa \approx \varepsilon^2 \operatorname{Ric}/N$, it fails by a factor $1/\varepsilon$ compared to the Bonnet–Myers theorem!), but is sharp in many other examples.

For the discrete cube $X = \{0,1\}^N$ (Example 8 above), one has $J = 1/2$ and $\kappa = 1/N$, so we get $\operatorname{diam} X \leqslant N$ which is the exact value.

For the discrete Ornstein–Uhlenbeck process (Example 10 above) one has $J = 1/2$ and $\kappa = 1/2N$, so we get $\operatorname{diam} X \leqslant 2N$ which once more is the exact value.

For the continuous Ornstein–Uhlenbeck process on $\mathbb{R}$ (Example 9 with $N = 1$), the diameter is infinite, consistently with the fact that J is unbounded. If we consider points x, y lying in some large interval $[-R; R]$ with $R \gg s/\sqrt{\alpha}$, then $\sup J \sim \alpha R \delta t$ on this interval, and $\kappa = (1 - \mathrm{e}^{\alpha \delta t}) \sim \alpha \delta t$ so that the diameter bound is $2R$, which is correct.

These examples show that one cannot replace J/κ with $J/\sqrt{\kappa}$ in this result (as could be expected from the example of Riemannian manifolds). In fact, Riemannian manifolds seem to be the only simple example where there is a diameter bound behaving like $1/\sqrt{\kappa}$. In Section 6 we investigate conditions under which an L^2 version of the Bonnet–Myers theorem holds.

In case J is not bounded, we can estimate instead the “average” diameter $\int d(x,y)\,\mathrm{d}\nu(x)\mathrm{d}\nu(y)$ under the invariant distribution ν. This estimate will prove very useful in several examples, to get bounds on the average of $\sigma(x)$ in cases where $\sigma(x)$ is unbounded but controlled by the distance to some “origin” (see e.g. Sections 3.3.4 and 3.3.5).

Proposition 24 (Average L^1 Bonnet–Myers).
Suppose that $\kappa(x,y) \geqslant \kappa > 0$ for any two distinct $x, y \in X$. Then for any $x \in X$,

$$\int_X d(x,y)\,\mathrm{d}\nu(y) \leqslant \frac{J(x)}{\kappa}$$

and so

$$\int_{X\times X} d(x,y)\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(y) \leqslant \frac{2\inf_x J(x)}{\kappa}$$

Proof.
The first assertion follows from Corollary 22 with $f = d(x,\cdot)$.

For the second assertion, choose an x_0 with $J(x_0)$ arbitrarily close to $\inf J$, and write

$$\begin{aligned}\int_{X\times X} d(y,z)\,\mathrm{d}\nu(y)\,\mathrm{d}\nu(z) &\leqslant \int_{X\times X} (d(y,x_0)+d(x_0,z))\,\mathrm{d}\nu(y)\,\mathrm{d}\nu(z)\\ &= 2W_1(\delta_{x_0},\nu) \leqslant 2J(x_0)/\kappa\end{aligned}$$

which ends the proof. □

2.4 Three constructions

Here we describe three very simple operations which trivially preserve positive curvature, namely, composition, superposition and L^1 tensorization.

Proposition 25 (Composition).
Let X be a metric space equipped with two random walks $m = (m_x)_{x\in X}$, $m' = (m'_x)_{x\in X}$. Suppose that the coarse Ricci curvature of m (resp. m') is at least κ (resp. κ'). Let m'' be the composition of m and m', i.e. the random walk which sends a probability measure μ to $\mu * m * m'$. Then the coarse Ricci curvature of m'' is at least $\kappa+\kappa'-\kappa\kappa'$.

Proof.
Trivial when $(1-\kappa)$ is seen as a contraction coefficient. □

Superposition states that if we are given two random walks on the same space and construct a new one by, at each step, tossing a coin and deciding to follow either one random walk or the other, then the coarse Ricci curvatures mix nicely.

Proposition 26 (Superposition).
Let X be a metric space equipped with a family $(m^{(i)})$ of random walks. Suppose that for each i, the coarse Ricci curvature of $m^{(i)}$ is at least κ_i. Let (α_i) be a family of non-negative real numbers with $\sum\alpha_i = 1$. Define a random walk m on X by $m_x := \sum \alpha_i m_x^{(i)}$. Then the coarse Ricci curvature of m is at least $\sum\alpha_i\kappa_i$.

Proof.
Let $x,y\in X$ and for each i let ξ_i be a coupling between $m_x^{(i)}$ and $m_y^{(i)}$. Then $\sum\alpha_i\xi_i$ is a coupling between $\sum\alpha_i m_x^{(i)}$ and $\sum\alpha_i m_y^{(i)}$, so that

$$\begin{aligned}W_1(m_x,m_y) &\leqslant \sum \alpha_i\, W_1\left(m_x^{(i)}, m_y^{(i)}\right)\\ &\leqslant \sum \alpha_i(1-\kappa_i)d(x,y)\\ &= \left(1-\textstyle\sum\alpha_i\kappa_i\right)d(x,y)\end{aligned}$$

Note that the coupling above, which consists in sending each $m_x^{(i)}$ to $m_y^{(i)}$, has no reason to be optimal, so that in general equality does not hold. □

Tensorization states that if we perform a random walk in a product space by deciding at random, at each step, to move in one or the other component, then positive curvature is preserved.

Proposition 27 (L^1 tensorization).
Let $((X_i, d_i))_{i\in I}$ be a finite family of metric spaces and suppose that X_i is equipped with a random walk $m^{(i)}$. Let X be the product of the spaces X_i, equipped with the distance $d := \sum d_i$. Let (α_i) be a family of non-negative real numbers with $\sum \alpha_i = 1$. Consider the random walk m on X defined by

$$m_{(x_1,\ldots,x_k)} := \sum \alpha_i \; \delta_{x_1} \otimes \cdots \otimes m_{x_i} \otimes \cdots \otimes \delta_{x_k}$$

Suppose that for each i, the coarse Ricci curvature of $m^{(i)}$ is at least κ_i. Then the coarse Ricci curvature of m is at least $\inf \alpha_i\kappa_i$.

For example, this allows for a very short proof that the curvature of the lazy random walk on the discrete cube $\{0,1\}^N$ is $1/N$ (Example 8). Indeed, it is the N-fold product of the random walk on $\{0,1\}$ which sends each point to the equilibrium distribution $(1/2,1/2)$, hence is of curvature 1.

Likewise, we can recover the coarse Ricci curvature for multinomial distributions (Example 12) as follows: Consider a finite set S of cardinal $d+1$, representing the boxes of Example 12, endowed with an arbitrary probability distribution ν. Equip it with the trivial distance and the Markov chain sending each point of S to ν, so that coarse Ricci curvature is 1. Now consider the N-fold product of this random walk on S^N. Each component represents a ball of Example 12, and the product random walk consists in selecting a ball and putting it in a random box according to ν, as in the example. By the proposition above, the coarse Ricci curvature of this N-fold product is (at least) $1/N$. This evaluation of curvature carries down to the "quotient" Markov chain of Example 12, in which only the number of balls in each box is considered instead of the full configuration space.

The case when some α_i is equal to 0 shows why coarse Ricci curvature is given by an infimum: indeed, if $\alpha_i = 0$ then the corresponding component never gets mixed, hence curvature cannot be positive (unless this component is reduced to a single point). This is similar to what happens for the spectral gap.

The statement above is restricted to a finite product for the following technical reasons: First, to define the L^1 product of an infinite family, a basepoint has to be chosen. Second, in order for the formula above to define a random walk with finite first moment (see Definition 1), some uniform assumption on the first moments of the $m^{(i)}$ is needed.

Proof.
For $x \in X$ let $\tilde m_x^{(i)}$ stand for $\delta_{x_1} \otimes \cdots \otimes m_{x_i} \otimes \cdots \otimes \delta_{x_k}$.

Let $x = (x_i)$ and $y = (y_i)$ be two points in X. Then

$$\begin{aligned}
W_1(m_x, m_y) &\leqslant \sum \alpha_i \, W_1\left(\tilde{m}_x^{(i)}, \tilde{m}_y^{(i)}\right) \\
&\leqslant \sum \alpha_i \left(W_1\left(m_x^{(i)}, m_y^{(i)}\right) + \sum_{j\neq i} d_j(x_j, y_j) \right) \\
&\leqslant \sum \alpha_i \left((1-\kappa_i) d_i(x_i, y_i) + \sum_{j\neq i} d_j(x_j, y_j) \right) \\
&= \sum \alpha_i \left(-\kappa_i d_i(x_i, y_i) + \sum d_j(x_j, y_j) \right) \\
&= \sum d_i(x_i, y_i) - \sum \alpha_i \kappa_i d_i(x_i, y_i) \\
&\leqslant (1 - \inf \alpha_i \kappa_i) \sum d_i(x_i, y_i) \\
&= (1 - \inf \alpha_i \kappa_i) \, d(x, y)
\end{aligned}$$

□

2.5 Lipschitz functions and spectral gap

Definition 28 (Averaging operator, Laplacian).
For $f \in L^2(X, \nu)$ let the averaging operator M *be*

$$\mathrm{M}f(x) := \int_y f(y) \, \mathrm{d}m_x(y)$$

and let $\Delta := \mathrm{M} - \mathrm{Id}$.

(This is the layman's convention for the sign of the Laplacian, i.e. $\Delta = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$ on $\mathbb{R}$, so that on a Riemannian manifold Δ is a negative operator.)

The following proposition also appears in [DGW04] (in the proof of Proposition 2.10). For the classical case of Riemannian manifolds, contraction of the norm of the gradient is one of the main results of Bakry–Émery theory.

Proposition 29 (Lipschitz contraction).
Let (X, d, m) be a random walk on a metric space. Let $\kappa \in \mathbb{R}$.

Then the coarse Ricci curvature of X is at least κ, if and only if, for every k-Lipschitz function $f : X \to \mathbb{R}$, the function $\mathrm{M}f$ is $k(1-\kappa)$-Lipschitz.

Proof.
First, suppose that the coarse Ricci curvature of X is at least κ. Then, using the

notation presented at the end of Section 1.1, we have

$$\begin{aligned}
\mathrm{M}f(y) - \mathrm{M}f(x) &= \int_z f(y+z) - f(x+z) \\
&\leqslant k \int_z d(x+z, y+z) \\
&= k\, d(x,y)(1-\kappa(x,y))
\end{aligned}$$

Conversely, suppose that whenever f is 1-Lipschitz, $\mathrm{M}f$ is $(1-\kappa)$-Lipschitz. The duality theorem for transportation distance (Theorem 1.14 in [Vil03]) states that

$$\begin{aligned}
W_1(m_x, m_y) &= \sup_{f \text{ 1-Lipschitz}} \int f \,\mathrm{d}(m_x - m_y) \\
&= \sup_{f \text{ 1-Lipschitz}} \mathrm{M}f(x) - \mathrm{M}f(y) \\
&\leqslant (1-\kappa) d(x,y)
\end{aligned}$$

□

Let ν be an invariant distribution of the random walk. Consider the space $L^2(X,\nu)/\{\mathrm{const}\}$ equipped with the norm

$$\|f\|^2_{L^2(X,\nu)/\{\mathrm{const}\}} := \|f - \mathbb{E}_\nu f\|^2_{L^2(X,\nu)} = \mathrm{Var}_\nu f = \frac{1}{2}\int_{X\times X} (f(x)-f(y))^2 \,\mathrm{d}\nu(x)\,\mathrm{d}\nu(y)$$

The operators M and Δ are self-adjoint in $L^2(X,\nu)$ if and only if ν is reversible for the random walk.

It is easy to check, using associativity of variances, that

$$\mathrm{Var}_\nu f = \int \mathrm{Var}_{m_x} f \,\mathrm{d}\nu(x) + \mathrm{Var}_\nu \mathrm{M}f$$

so that $\|\mathrm{M}f\|_2 \leqslant \|f\|_2$. It is also clear that $\|\mathrm{M}f\|_\infty \leqslant \|f\|_\infty$.

Usually, spectral gap properties for Δ are expressed in the space L^2. The proposition above only implies that the spectral radius of the operator M acting on $\mathrm{Lip}(X)/\{\mathrm{const}\}$ is at most $(1-\kappa)$. In general it is not true that a bound for the spectral radius of an operator on a dense subspace of a Hilbert space implies a bound for the spectral radius on the whole space. This holds, however, when the operator is self-adjoint or when the Hilbert space is finite-dimensional.

Proposition 30 (Spectral gap).
Let (X,d,m) be a metric space with random walk, with invariant distribution ν. Suppose that the coarse Ricci curvature of X is at least $\kappa > 0$ and that $\sigma < \infty$. Suppose that ν is reversible, or that X is finite.

Then the spectral radius of the averaging operator acting on $L^2(X,\nu)/\{\mathrm{const}\}$ is at most $1-\kappa$.

Compare Theorem 1.9 in [CW94] (Theorem 9.18 in [Che04]).

Proof.

First, if X is finite then Lipschitz functions coincide with L^2 functions, and the norms are equivalent, so that there is nothing to prove. So we suppose that ν is reversible, i.e. M is self-adjoint.

Let f be a k-Lipschitz function. Proposition 32 below implies that Lipschitz functions belong to $L^2(X,\nu)/\{\text{const}\}$ and that the Lipschitz norm controls the L^2 norm (this is where we use that $\sigma < \infty$). Since $\mathrm{M}^t f$ is $k(1-\kappa)^t$-Lipschitz one gets $\operatorname{Var} \mathrm{M}^t f \leqslant Ck^2(1-\kappa)^{2t}$ for some constant C so that $\lim_{t\to\infty}(\sqrt{\operatorname{Var} \mathrm{M}^t f})^{1/t} \leqslant (1-\kappa)$. So the spectral radius of M is at most $1-\kappa$ on the subspace of Lipschitz functions.

Now Lipschitz functions are dense in $L^2(X,\nu)$ (indeed, a probability measure on a metric space is regular, so that indicator functions of measurable sets can be approximated by Lipschitz functions). Since M is bounded and self-adjoint, its spectral radius is controlled by its value on a dense subspace using the spectral decomposition. □

Corollary 31 (Poincaré inequality).

Let (X, d, m) be an ergodic random walk on a metric space, with invariant distribution ν. Suppose that the coarse Ricci curvature of X is at least $\kappa > 0$ and that $\sigma < \infty$. Suppose that ν is reversible.

Then the spectrum of $-\Delta$ acting on $L^2(X,\nu)/\{\text{const}\}$ is contained in $[\kappa;\infty)$. Moreover the following discrete Poincaré inequalities are satisfied for $f \in L^2(X,\nu)$:

$$\operatorname{Var}_\nu f \leqslant \frac{1}{\kappa(2-\kappa)} \int \operatorname{Var}_{m_x} f \, \mathrm{d}\nu(x)$$

and

$$\operatorname{Var}_\nu f \leqslant \frac{1}{2\kappa} \iint (f(y)-f(x))^2 \, \mathrm{d}\nu(x) \, \mathrm{d}m_x(y)$$

Proof.

These are rewritings of the inequalities $\operatorname{Var}_\nu \mathrm{M}f \leqslant (1-\kappa)^2 \operatorname{Var}_\nu f$ and $\langle f, \mathrm{M}f\rangle_{L^2(X,\nu)/\{\text{const}\}} \leqslant (1-\kappa) \operatorname{Var}_\nu f$, respectively. □

The quantities $\operatorname{Var}_{m_x} f$ and $\frac{1}{2}\int (f(y)-f(x))^2 \, \mathrm{d}m_x(y)$ are two possible analogues of $\|\nabla f(x)\|^2$ in a discrete setting. Though the latter is more common, the former is preferable when the support of m_x can be far away from x because it cancels out the "drift". Moreover one always has $\operatorname{Var}_{m_x} f \leqslant \int (f(y)-f(x))^2 \, \mathrm{d}m_x(y)$, so that the first form is generally sharper.

Reversibility is really needed here to turn an estimate of the spectral radius of M into an inequality between the norms of Mf and f, using that M is self-adjoint. When the random walk is not reversible, applying the above to MM* does not work since the coarse Ricci curvature of the latter is unknown. However, a version of the Poincaré inequality with a non-local gradient still holds (Theorem 45).

As proven by Gromov and Milman ([GM83], or Corollary 3.1 and Theorem 3.3 in [Led01]), in quite a general setting a Poincaré inequality implies exponential concentration. Their argument adapts well here, and provides a concentration bound of roughly $\exp(-t\sqrt{\kappa}\,\sigma_\infty)$. We do not include the details, however, since Theorem 33 below is always more precise and covers the non-reversible case as well.

Let us compare this result to Lichnerowicz's theorem in the case of the ε-step random walk on an N-dimensional Riemannian manifold with positive Ricci curvature. This theorem states that the smallest eigenvalue of the usual Laplacian is $\frac{N}{N-1}\inf \mathrm{Ric}$, where $\inf \mathrm{Ric}$ is the largest K such that $\mathrm{Ric}(v, v) \geqslant K$ for all unit tangent vectors v. On the other hand, the operator Δ associated with the random walk is the difference between the mean value of a function on a ball of radius ε, and its value at the center of the ball: when $\varepsilon \to 0$ this behaves like $\frac{\varepsilon^2}{2(N+2)}$ times the usual Laplacian (take the average on the ball of the Taylor expansion of f). We saw (Example 7) that in this case $\kappa \sim \frac{\varepsilon^2}{2(N+2)}\inf \mathrm{Ric}$. Note that both scaling factors are the same. So we miss the $\frac{N}{N-1}$ factor, but otherwise get the correct order of magnitude.

Second, let us test this corollary for the discrete cube of Example 8. In this case the eigenbase of the discrete Laplacian is well-known (characters, or Fourier/Walsh transform), and the spectral gap of the discrete Laplacian associated with the lazy random walk is $1/N$. Since the coarse Ricci curvature κ is $1/N$ too, the value given in the proposition is sharp.

Third, consider the Ornstein–Uhlenbeck process on $\mathbb{R}$, as in Example 9. Its infinitesimal generator is $L = \frac{s^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} - \alpha x \frac{\mathrm{d}}{\mathrm{d}x}$, and the eigenfunctions are known to be $H_k(x\sqrt{\alpha/s^2})$ where H_k is the Hermite polynomial $H_k(x) := (-1)^k \mathrm{e}^{x^2}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\mathrm{e}^{-x^2}$. The associated eigenvalue of L is $-k\alpha$, so that the spectral gap of L is α. Now the random walk we consider is the flow $\mathrm{e}^{\delta t L}$ at time δt of the process (with small δt), whose eigenvalues are $\mathrm{e}^{-k\alpha\delta t}$. So the spectral gap of the discrete Laplacian $\mathrm{e}^{\delta t L} - \mathrm{Id}$ is $1 - \mathrm{e}^{-\alpha\delta t}$. Since coarse Ricci curvature is $1 - \mathrm{e}^{-\alpha\delta t}$ too, the corollary is sharp again.

3 Concentration results

3.1 Variance of Lipschitz functions

We begin with the simplest kind of concentration, namely, an estimation of the variance of Lipschitz functions. Contrary to Gaussian or exponential concentration, the only assumption needed here is that the average diffusion constant σ is finite.

Since our Gaussian concentration result will yield basically the same variance $\sigma^2/n\kappa$, we discuss sharpness of this estimate in various examples in Section 3.3.

Proposition 32.
Let (X, d, m) be a random walk on a metric space, with coarse Ricci curvature at least $\kappa > 0$. Let ν be the unique invariant distribution. Suppose that $\sigma < \infty$.

Then the variance of a 1-Lipschitz function is at most $\frac{\sigma^2}{n\kappa(2-\kappa)}$.

Note that since $\kappa \leqslant 1$ one has $\frac{\sigma^2}{n\kappa(2-\kappa)} \leqslant \frac{\sigma^2}{n\kappa}$.

In particular, this implies that all Lipschitz functions are in $L^2/\{\mathrm{const}\}$; especially, $\int d(x, y)^2 \mathrm{d}\nu(x)\mathrm{d}\nu(y)$ is finite. The fact that the Lipschitz norm controls the L^2 norm was used above in the discussion of spectral properties of the random walk operator.

The assumption $\sigma < \infty$ is necessary. As a counterexample, consider a random walk on $\mathbb{N}$ that sends every $x \in \mathbb{N}$ to some fixed distribution ν on $\mathbb{N}$ with infinite second moment: coarse Ricci curvature is 1, yet the identity function is not in L^2.

Proof.
Suppose for now that $|f|$ is bounded by $A \in \mathbb{R}$, so that $\operatorname{Var} f < \infty$. We first prove that $\operatorname{Var} \mathrm{M}^t f$ tends to 0. Let B_r be the ball of radius r in X centered at some basepoint. Using that $\mathrm{M}^t f$ is $(1-\kappa)^t$-Lipschitz on B_r and bounded by A on $X \setminus B_r$, we get $\operatorname{Var} \mathrm{M}^t f = \frac{1}{2} \iint (\mathrm{M}^t f(x) - \mathrm{M}^t f(y))^2 \,\mathrm{d}\nu(x)\mathrm{d}\nu(y) \leqslant 2(1-\kappa)^{2t} r^2 + 2A^2 \nu(X \setminus B_r)$. Taking for example $r = 1/(1-\kappa)^{t/2}$ shows that $\operatorname{Var} \mathrm{M}^t f \to 0$.

As already mentioned, one has $\operatorname{Var} f = \operatorname{Var} \mathrm{M} f + \int \operatorname{Var}_{m_x} f \,\mathrm{d}\nu(x)$. Since $\operatorname{Var} \mathrm{M}^t f \to 0$, by induction we get

$$\operatorname{Var} f = \sum_{t=0}^{\infty} \int \operatorname{Var}_{m_x} \mathrm{M}^t f \,\mathrm{d}\nu(x)$$

Now since f is 1-Lipschitz, by definition $\operatorname{Var}_{m_x} f \leqslant \sigma(x)^2/n_x$. Since $\mathrm{M}^t f$ is $(1-\kappa)^t$-Lipschitz, we have $\operatorname{Var}_{m_x} \mathrm{M}^t f \leqslant (1-\kappa)^{2t}\, \sigma(x)^2/n_x$ so that the sum above is at most $\frac{\sigma^2}{n\kappa(2-\kappa)}$. The case of unbounded f is treated by a simple limiting argument. □

3.2 Gaussian concentration

As mentioned above, positive coarse Ricci curvature implies a Gaussian-then-exponential concentration theorem. The estimated variance is $\sigma^2/n\kappa$ as above, so that this is essentially a more precise version of Proposition 32, with some loss in the constants. We will see in the discussion below (Section 3.3) that in the main examples, the order of magnitude is correct.

The fact that concentration is not always Gaussian far away from the mean is genuine, as exemplified by binomial distributions on the cube (Section 3.3.3) or $M/M/\infty$ queues (Section 3.3.4). The width of the Gaussian window is controlled by two factors. First, variations of the diffusion constant $\sigma(x)^2$ can result in purely exponential behavior (Section 3.3.5); this leads to the assumption that $\sigma(x)^2$ is bounded by a Lipschitz function. Second, as Gaussian phenomena only emerge as the result of a large number of small events, the "granularity" of the process must be bounded, which leads to the (comfortable) assumption that $\sigma_\infty < \infty$. Otherwise, a Markov chain which sends every point $x \in X$ to some fixed measure ν has coarse Ricci curvature 1 and can have arbitrary bad concentration properties depending on ν.

In the case of Riemannian manifolds, simply letting the step of the random walk tend to 0 makes the width of the Gaussian window tend to infinity, so that we recover Gaussian concentration as in the Lévy–Gromov or Bakry–Émery theorems. For the uniform measure on the discrete cube, the Gaussian width is equal to the diameter of the cube, so that we get full Gaussian concentration as well. In a series of other examples (such as Poisson measures), the transition from Gaussian to non-Gaussian regime occurs roughly as predicted by the theorem.

Theorem 33 (Gaussian concentration).
Let (X, d, m) be a random walk on a metric space, with coarse Ricci curvature at least $\kappa > 0$. Let ν be the unique invariant distribution.

Let

$$D_x^2 := \frac{\sigma(x)^2}{n_x \kappa}$$

and

$$D^2 := \mathbb{E}_\nu D_x^2$$

Suppose that the function $x \mapsto D_x^2$ is C-Lipschitz. Set

$$t_{\max} := \frac{D^2}{\max(\sigma_\infty, 2C/3)}$$

Then for any 1-Lipschitz function f, for any $t \leqslant t_{\max}$ we have

$$\nu\left(\{x, f(x) \geqslant t + \mathbb{E}_\nu f\}\right) \leqslant \exp -\frac{t^2}{6D^2}$$

and for $t \geqslant t_{\max}$

$$\nu\left(\{x, f(x) \geqslant t + \mathbb{E}_\nu f\}\right) \leqslant \exp\left(-\frac{t_{\max}^2}{6D^2} - \frac{t - t_{\max}}{\max(3\sigma_\infty, 2C)}\right)$$

Remark 34.
Proposition 24 or Corollary 22 often provide very sharp a priori bounds for $\mathbb{E}_\nu D_x^2$ even when no information on ν is available, as we shall see in the examples.

Remark 35.
It is clear from the proof below that $\sigma(x)^2/n_x\kappa$ itself need not be Lipschitz, only bounded by some Lipschitz function. In particular, if $\sigma(x)^2$ is bounded one can always set $D^2 = \sup_x \frac{\sigma(x)^2}{n_x\kappa}$ and $C = 0$.

Remark 36 (Continuous-time situations).
If we replace the random walk $m = (m_x)_{x\in X}$ with the lazy random walk m' whose transition probabilities are $m'_x := (1-\alpha)\delta_x + \alpha m_x$, when α tends to 0 this approximates the law at time α of the continuous-time random walk with transition rates m_x, so that the continuous-time random walk is obtained by taking the lazy random walk m' and speeding up time by $1/\alpha$ when $\alpha \to 0$. Of course this does not change the invariant distribution. The point is that when $\alpha \to 0$, both σ_x^2 and κ scale like α (and n_x tends to 1), so that D^2 has a finite limit. This means that we can apply Theorem 33 to continuous-time examples that naturally appear as limits of a discrete-time, finite-space Markov chain, as illustrated in Sections 3.3.4 to 3.3.6.

Remark 37.
The condition that σ_∞ is uniformly bounded can be replaced with a Gaussian-type

assumption, namely that for each measure m_x there exists a number s_x such that $\mathbb{E}_{m_x} \mathrm{e}^{\lambda f} \leqslant \mathrm{e}^{\lambda^2 s_x^2/2} \mathrm{e}^{\lambda \mathbb{E}_{m_x} f}$ for any 1-Lipschitz function f. Then a similar theorem holds, with $\sigma(x)^2/n_x$ replaced with s_x^2. (When s_x^2 is constant this is Proposition 2.10 in [DGW04].) However, this is generally not well-suited to discrete settings, because when transition probabilities are small, the best s_x^2 for which such an inequality is satisfied is usually much larger than the actual variance $\sigma(x)^2$: for example, if two points x and y are at distance 1 and $m_x(y) = \varepsilon$, s_x must satisfy $s_x^2 \geqslant 1/2 \ln(1/\varepsilon) \gg \varepsilon$. Thus making this assumption will provide extremely poor estimates of the variance D^2 when some transition probabilities are small (e.g. for binomial distributions on the discrete cube), and in particular, this cannot extend to the continuous-time limit.

In Section 3.3.5, we give a simple example where the Lipschitz constant of $\sigma(x)^2$ is large, resulting in exponential rather than Gaussian behavior. In Section 3.3.6 we give two examples of positively curved process with heavy tails: one in which $\sigma_\infty = 1$ but with non-Lipschitz growth of $\sigma(x)^2$, and one with $\sigma(x)^2 \leqslant 1$ but with unbounded $\sigma_\infty(x)$. These show that the assumptions cannot be relaxed.

Proof.
This proof is a variation on standard martingale methods for concentration (see e.g. Lemma 4.1 in [Led01], or [Sch01]).

Lemma 38.
Let $\varphi : X \to \mathbb{R}$ be an α-Lipschitz function with $\alpha \leqslant 1$. Assume $\lambda \leqslant 1/3\sigma_\infty$. Then for $x \in X$ we have

$$(\mathrm{M}\mathrm{e}^{\lambda\varphi})(x) \leqslant \mathrm{e}^{\lambda \mathrm{M}\varphi(x) + \lambda^2 \alpha^2 \frac{\sigma(x)^2}{n_x}}$$

Note that the classical Proposition 1.16 in [Led01] would yield $(\mathrm{M}\mathrm{e}^{\lambda\varphi})(x) \leqslant \mathrm{e}^{\lambda \mathrm{M}\varphi(x) + 2\lambda^2\alpha^2\sigma_\infty^2}$, which is too weak to provide reasonable variance estimates.

Proof of the lemma.
For any smooth function g and any real-valued random variable Y, a Taylor expansion with Lagrange remainder gives $\mathbb{E} g(Y) \leqslant g(\mathbb{E}Y) + \frac{1}{2}(\sup g'') \operatorname{Var} Y$. Applying this with $g(Y) = \mathrm{e}^{\lambda Y}$ we get

$$(\mathrm{M}\mathrm{e}^{\lambda\varphi})(x) = \mathbb{E}_{m_x} \mathrm{e}^{\lambda\varphi} \leqslant \mathrm{e}^{\lambda \mathrm{M}\varphi(x)} + \frac{\lambda^2}{2} \left(\sup_{\operatorname{Supp} m_x} \mathrm{e}^{\lambda\varphi} \right) \operatorname{Var}_{m_x} \varphi$$

and note that since $\operatorname{diam} \operatorname{Supp} m_x \leqslant 2\sigma_\infty$ and φ is 1-Lipschitz we have $\sup_{\operatorname{Supp} m_x} \varphi \leqslant \mathbb{E}_{m_x} \varphi + 2\sigma_\infty$, so that

$$(\mathrm{M}\mathrm{e}^{\lambda\varphi})(x) \leqslant \mathrm{e}^{\lambda \mathrm{M}\varphi(x)} + \frac{\lambda^2}{2} \mathrm{e}^{\lambda \mathrm{M}\varphi(x) + 2\lambda\sigma_\infty} \operatorname{Var}_{m_x} \varphi$$

Now, by definition we have $\operatorname{Var}_{m_x} \varphi \leqslant \alpha^2 \frac{\sigma(x)^2}{n_x}$. Moreover for $\lambda \leqslant 1/3\sigma_\infty$ we have $\mathrm{e}^{2\lambda\sigma_\infty} \leqslant 2$, hence the result. □

Back to the proof of the theorem, let f be a 1-Lipschitz function and $\lambda \geqslant 0$. Define by induction $f_0 := f$ and $f_{k+1}(x) := \mathrm{M} f_k(x) + \lambda \frac{\sigma(x)^2}{n_x}(1-\kappa/2)^{2k}$.

Suppose that $\lambda \leqslant 1/2C$. Then $\lambda\frac{\sigma(x)^2}{n_x}$ is $\kappa/2$-Lipschitz. Using Proposition 29, we can show by induction that f_k is $(1-\kappa/2)^k$-Lipschitz.

Consequently, the lemma yields

$$(\mathrm{M}\mathrm{e}^{\lambda f_k})(x) \leqslant \mathrm{e}^{\lambda \mathrm{M} f_k(x) + \lambda^2 \frac{\sigma(x)^2}{n_x}(1-\kappa/2)^{2k}} = \mathrm{e}^{\lambda f_{k+1}(x)}$$

so that by induction

$$(\mathrm{M}^k \mathrm{e}^{\lambda f})(x) \leqslant \mathrm{e}^{\lambda f_k(x)}$$

Now setting $g(x) := \frac{\sigma(x)^2}{n_x}$ and unwinding the definition of f_k yields

$$f_k(x) = (\mathrm{M}^k f)(x) + \lambda \sum_{i=1}^{k} (\mathrm{M}^{k-i} g)(x)\,(1-\kappa/2)^{2(i-1)}$$

so that

$$\lim_{k\to\infty} f_k(x) = \mathbb{E}_\nu f + \lambda \sum_{i=1}^{\infty} \mathbb{E}_\nu g\,(1-\kappa/2)^{2(i-1)} \leqslant \mathbb{E}_\nu f + \lambda \mathbb{E}_\nu g\,\frac{4}{3\kappa}$$

Meanwhile, $(\mathrm{M}^k \mathrm{e}^{\lambda f})(x)$ tends to $\mathbb{E}_\nu \mathrm{e}^{\lambda f}$, so that

$$\mathbb{E}_\nu \mathrm{e}^{\lambda f} \leqslant \lim_{k\to\infty} \mathrm{e}^{\lambda f_k} \leqslant \mathrm{e}^{\lambda \mathbb{E}_\nu f + \frac{4\lambda^2}{3\kappa}\mathbb{E}_\nu \frac{\sigma(x)^2}{n_x}}$$

We can conclude by a standard Chebyshev inequality argument. The restrictions $\lambda \leqslant 1/2C$ and $\lambda \leqslant 1/3\sigma_\infty$ give the value of $t_{\max}$. □

Remark 39 (Finite-time concentration).
The proof provides a similar concentration result for the finite-time measures m_x^{*k} as well, with variance

$$D_{x,k}^2 = \sum_{i=1}^{k} (1-\kappa/2)^{2(i-1)} \left(\mathrm{M}^{k-i}\frac{\sigma(y)^2}{n_y}\right)(x)$$

and $D_{x,k}^2$ instead of D^2 in the expression for $t_{\max}$.

3.3 Examples revisited

Let us test the sharpness of these estimates in some examples, beginning with the simplest ones. In each case, we gather the relevant quantities in a table. Recall that $\approx$ denotes an equality up to a multiplicative universal constant (typically $\leqslant 4$), while symbol $\sim$ denotes usual asymptotic equivalence (with sharp constant).

3.3.1 Riemannian manifolds

First, let X be a compact N-dimensional Riemannian manifold with positive Ricci curvature. Equip this manifold with the ε-step random walk as in Example 7. The measure $\frac{\operatorname{vol} B(x,\varepsilon)}{\operatorname{vol} B_{\mathrm{Eucl}}(\varepsilon)} \operatorname{dvol}(x)$ is reversible for this random walk. In particular, when $\varepsilon \to 0$, the density of this measure with respect to the Riemannian volume is $1+O(\varepsilon^2)$.

Let $\inf \operatorname{Ric}$ denote the largest $K > 0$ such that $\operatorname{Ric}(v,v) \geqslant K$ for any unit tangent vector v. The relevant quantities for this random walk are as follows (see Section 8 for the proofs).

Coarse Ricci curvature	$\kappa \sim \frac{\varepsilon^2}{2(N+2)} \inf \operatorname{Ric}$
Coarse diffusion constant	$\sigma(x)^2 \sim \varepsilon^2 \frac{N}{N+2} \quad \forall x$
Dimension	$n \approx N$
Variance (Lévy–Gromov thm.)	$\approx 1/\inf \operatorname{Ric}$
Gaussian variance (Thm. 33)	$D^2 \approx 1/\inf \operatorname{Ric}$
Gaussian range	$t_{\max} \approx 1/(\varepsilon \inf \operatorname{Ric}) \to \infty$

So, up to some (small) numerical constants, we recover Gaussian concentration as in the Lévy-Gromov theorem.

The same applies to diffusions with a drift on a Riemannian manifold, as considered by Bakry and Émery. To be consistent with the notation of Example 11, in the table above ε has to be replaced with $\sqrt{(N+2)\delta t}$, and $\inf \operatorname{Ric}$ with $\inf\left(\operatorname{Ric}(v,v) - 2\nabla^{\mathrm{sym}} F(v,v)\right)$ for v a unit tangent vector. (In the non-compact case, care has to be taken since the solution of the stochastic differential equation of Example 11 on the manifold may not exist, and even if it does its Euler scheme approximation at time δt may not converge uniformly on the manifold. In explicit examples such as the Ornstein–Uhlenbeck process, however, this is not a problem.)

3.3.2 Discrete cube

Consider now the discrete cube $\{0,1\}^N$ equipped with its graph distance (Hamming metric) and lazy random walk (Example 8).

For a random walk on a graph one always has $\sigma \approx 1$, and $n \geqslant 1$ in full generality. The following remark allows for more precise constants.

Remark 40.
Let m be a random walk on a graph. Then, for any vertex x one has $\sigma(x)^2/n_x \leqslant 1 - m_x(\{x\})$.

Proof.
By definition $\sigma(x)^2/n_x$ is the maximal variance, under m_x, of a 1-Lipschitz function. So let f be a 1-Lipschitz function on the graph. Since variance is invariant by adding a constant, we can assume that $f(x) = 0$. Then $|f(y)| \leqslant 1$ for any neighbor y of x. The mass, under m_x, of all neighbors of x is $1 - m_x(\{x\})$. Hence $\operatorname{Var}_{m_x} f = \mathbb{E}_{m_x} f^2 - (\mathbb{E}_{m_x} f)^2 \leqslant \mathbb{E}_{m_x} f^2 \leqslant 1 - m_x(\{x\})$.

This value is achieved, for example, with a lazy simple random walk when x has an even number of neighbors and when no two distinct neighbors of x are mutual neighbors; in this case one can take $f(x) = 0$, $f = 1$ on half the neighbors of x and $f = -1$ on the remaining neighbors of x. □

Applying this to the lazy random walk on the discrete cube, one gets:

Coarse Ricci curvature	$\kappa = 1/N$
Coarse diffusion constant & dimension	$\sigma(x)^2/n_x \leqslant 1/2$
Estimated variance (Prop. 32)	$\sigma^2/n\kappa(2-\kappa) \sim N/4$
Actual variance	$N/4$
Gaussian variance (Thm. 33)	$D^2 \leqslant N/2$
Gaussian range	$t_{\max} = N/2$

In particular, since $N/2$ is the maximal possible value for the deviation from average of a 1-Lipschitz function on the cube, we see that $t_{\max}$ has the largest possible value.

3.3.3 Binomial distributions

The occurrence of a finite range $t_{\max}$ for the Gaussian behavior of tails is genuine, as the following example shows.

Let again $X = \{0,1\}^N$ equipped with its Hamming metric (each edge is of length 1). Consider the following Markov chain on X: for some $0 < p < 1$, at each step, choose a bit at random among the N bits; if it is equal to 0, flip it to 1 with probability p; if it is equal to 1, flip it to 0 with probability $1-p$. The binomial distribution $\nu\left((x_1,\ldots,x_N)\right) = \prod p^{x_i}(1-p)^{1-x_i}$ is reversible for this Markov chain.

The coarse Ricci curvature of this Markov chain is $1/N$, as can easily be seen directly or using the tensorization property (Proposition 27).

Let k be the number of bits of $x \in X$ which are equal to 1. Then k follows a Markov chain on $\{0,1,\ldots,N\}$, whose transition probabilities are:

$$\begin{aligned} p_{k,k+1} &= p(1-k/N) \\ p_{k,k-1} &= (1-p)k/N \\ p_{k,k} &= pk/N + (1-p)(1-k/N) \end{aligned}$$

The binomial distribution with parameters N and p, namely $\binom{N}{k}p^k(1-p)^{N-k}$, is reversible for this Markov chain. Moreover, the coarse Ricci curvature of this "quotient" Markov chain is still $1/N$.

Now, fix some $\lambda > 0$ and consider the case $p = \lambda/N$. Let $N \to \infty$. It is well-known that the invariant distribution tends to the Poisson distribution $\mathrm{e}^{-\lambda}\lambda^k/k!$ on $\mathbb{N}$.

Let us see how Theorem 33 performs on this example. The table below applies either to the full space $\{0,1\}^N$, with k the function "number of 1's", or to its projection on $\{0,1,\ldots,N\}$. Note the use of Proposition 24 to estimate σ^2 a priori, without having to resort to explicit knowledge of the invariant distribution. All constants implied in the $O(1/N)$ notation are small and completely explicit.

Coarse Ricci curvature	$\kappa = 1/N$
Coarse diffusion constant	$\sigma(k)^2 = (\lambda + k)/N + O(1/N^2)$
Estimated $\mathbb{E}k$ (Prop. 24)	$\mathbb{E}k \leqslant J(0)/\kappa = \lambda$
Actual $\mathbb{E}k$	$\mathbb{E}k = \lambda$
Average diffusion constant	$\sigma^2 = \mathbb{E}\sigma(k)^2 = 2\lambda/N + O(1/N^2)$
Dimension	$n \geqslant 1$
Estimated variance (Prop. 32)	$\sigma^2/n\kappa(2-\kappa) \leqslant \lambda + O(1/N)$
Actual variance	λ
Gaussian variance (Thm. 33)	$D^2 \leqslant 2\lambda + O(1/N)$
Lipschitz constant of D_x^2	$C = 1 + O(1/N)$
Gaussian range	$t_{\max} = 4\lambda/3$

The Poisson distribution has a roughly Gaussian behavior (with variance λ) in a range of size approximately λ around the mean; further away, it decreases like $\mathrm{e}^{-k \ln k}$ which is not Gaussian. This is in good accordance with $t_{\max}$ the table above, and shows that the Gaussian range cannot be extended.

3.3.4 A continuous-time example: $M/M/\infty$ queues

Here we show how to apply Theorem 33 to a continuous-time example, the $M/M/\infty$ queue. These queues were brought to my attention by D. Chafaï.

The $M/M/\infty$ queue consists of an infinite number of "servers". Each server can be free (0) or busy (1). The state space consists of all sequences in $\{0,1\}^{\mathbb{N}}$ with a finite number of 1's. The dynamics is at follows: Fix two numbers $\lambda > 0$ and $\mu > 0$. At a rate λ per unit of time, a client arrives and the first free server becomes busy. At a rate μ per unit of time, each busy server finishes its job (independently of the others) and becomes free. The number $k \in \mathbb{N}$ of busy servers is a continuous-time Markov chain, whose transition probabilities at small times t are given by

$$\begin{aligned} p^t_{k,k+1} &= \lambda t + O(t^2) \\ p^t_{k,k-1} &= k\mu t + O(t^2) \\ p^t_{k,k} &= 1 - (\lambda + k\mu)t + O(t^2) \end{aligned}$$

This system is often presented as a discrete analogue of an Ornstein–Uhlenbeck process, since asymptotically the drift is linear towards the origin. However, it is not symmetric around the mean, and moreover the invariant (actually reversible) distribution ν is a Poisson distribution with parameter λ/μ, rather than a Gaussian.

This continuous-time Markov chain can be seen as a limit of the binomial Markov chain above as follows: First, replace the binomial Markov chain with its continuous-time equivalent (Remark 36); Then, set $p = \lambda/N$ and let $N \to \infty$, while speeding up time by a factor N. The analogy is especially clear in the table below: if we replace λ with λ/N and μ with $1/N$, we get essentially the same table as for the binomial distribution.

It is easy to check that Proposition 32 (with $\sigma^2/2n\kappa$ instead of $\sigma^2/n\kappa(2-\kappa)$) and Theorem 33 pass to the limit. In this continuous-time setting, the definitions become the following: $\kappa(x,y) := -\frac{\mathrm{d}}{\mathrm{d}t} W_1(m_x^t, m_y^t)/d(x,y)$ (as mentioned in the introduction) and $\sigma(x)^2 := \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\iint d(y,z)\,\mathrm{d}m_x^t(y)\mathrm{d}m_x^t(z)$, where m_x^t is the law at time t of the process starting at x.

Then the relevant quantities are as follows.

Coarse Ricci curvature	$\kappa = \mu$
Coarse diffusion constant	$\sigma(k)^2 = k\mu + \lambda$
Estimated $\mathbb{E}k$ (Prop. 24)	$\mathbb{E}k \leqslant J(0)/\kappa = \lambda/\mu$
Actual $\mathbb{E}k$	$\mathbb{E}k = \lambda/\mu$
Average diffusion constant	$\sigma^2 = \mathbb{E}\sigma(k)^2 = 2\lambda$
Dimension	$n \geqslant 1$
Estimated variance (Prop. 32)	$\sigma^2/2n\kappa = \lambda/\mu$
Actual variance	λ/μ
Gaussian variance (Thm. 33)	$D^2 \leqslant 2\lambda/\mu$
Lipschitz constant of D_x^2	$C = 1$
Gaussian range	$t_{\max} = 4\lambda/3\mu$

So once more Theorem 33 is in good accordance with the behavior of the random walk, whose invariant distribution is Poisson with mean and variance λ/μ, thus Gaussian-like only in some interval around this value.

An advantage of this approach is that is can be generalized to situations where the rates of the servers are not constant, but, say, bounded between $\mu_0/10$ and $10\mu_0$, and clients go to the first free server according to some predetermined scheme, e.g. the fastest free server. Indeed, the $M/M/\infty$ queue above can be seen as a Markov chain in the full configuration space of the servers, namely the space of all sequences over the alphabet $\{\text{free}, \text{busy}\}$ containing a finite number of “busy”. It is easy to check that the coarse Ricci curvature is still equal to μ in this configuration space. Now in the case of variable rates, the number of busy servers is generally not Markovian, so one has to work in the configuration space. If the rate of the i-th server is μ_i, the coarse Ricci curvature is $\inf \mu_i$ in the configuration space, whereas the diffusion constant is controlled by $\sup \mu_i$. So if the rates vary in a bounded range, coarse Ricci curvature still provides a Gaussian-type control, though an explicit description of the invariant distribution is not available.

Let us consider more realistic queue models, such as the $M/M/k$ queue, i.e. the number of servers is equal to k (with constant or variable rates). Then, on the part of the space where some servers are free, coarse Ricci curvature is at least equal to the rate of the slowest server; whereas on the part of the space where all servers are busy, coarse Ricci curvature is 0. If, as often, an abandon rate for waiting clients is added to the model, then coarse Ricci curvature is equal to this abandon rate on the part of the space where all servers are busy (and in particular, coarse Ricci curvature is positive on the whole space).

3.3.5 An example of exponential concentration

We give here a very simple example of a Markov chain which has positive curvature but for which concentration is not Gaussian but exponential, due to large variations of the diffusion constant, resulting in a large value of C. Compare Example 14 above where exponential concentration was due to unbounded σ_∞.

This is a continuous-time random walk on $\mathbb{N}$ defined as follows. Take $0 < \alpha < \beta$. For $k \in \mathbb{N}$, the transition rate from k to $k+1$ is $(k+1)\alpha$, whereas the transition rate from $k+1$ to k is $(k+1)\beta$. It is immediate to check that the geometric distribution with decay α/β is reversible for this Markov chain.

The coarse Ricci curvature of this Markov chain is easily seen to be $\beta-\alpha$. We have $\sigma(k)^2 = (k+1)\alpha + k\beta$, so that $\sigma(k)^2$ is $(\alpha+\beta)$-Lipschitz and $C = (\alpha+\beta)/(\beta-\alpha)$.

The expectation of k under the invariant distribution can be bounded by $J(0)/\kappa = \alpha/(\beta-\alpha)$ by Proposition 24, which is actually the exact value. So the expression above for $\sigma(k)^2$ yields $\sigma^2 = 2\alpha\beta/(\beta-\alpha)$. Consequently, the estimated variance $\sigma^2/2n\kappa$ (obtained by the continuous-time version of Proposition 32) is at most $\alpha\beta/(\beta-\alpha)^2$, which is the actual value.

Now consider the case when $\beta - \alpha$ is small. If the C factor in Theorem 33 is not taken into account, we get blatantly false results since the invariant distribution is not Gaussian at all. Indeed, in the regime where $\beta - \alpha \to 0$, the width of the Gaussian window in Theorem 33 is $D^2/C \approx \alpha/(\beta-\alpha)$. This is fine, as this is the decay distance of the invariant distribution, and in this interval both the Gaussian and geometric estimates are close to 1 anyway. But without the C factor, we would get $D^2/\sigma_\infty = \alpha\beta/(\beta-\alpha)^2$, which is much larger; the invariant distribution is clearly not Gaussian on this interval.

Moreover, Theorem 33 predicts, in the exponential regime, a $\exp(-t/2C)$ behavior for concentration. Here the asymptotic behavior of the invariant distribution is $(\alpha/\beta)^t \sim (1-2/C)^t \sim \mathrm{e}^{-2t/C}$ when $\beta - \alpha$ is small. So we see that (up to a constant 4) the exponential decay rate predicted by Theorem 33 is genuine.

3.3.6 Heavy tails

It is clear that a variance control alone does not imply any concentration bound beyond the Bienaymé–Chebyshev inequality. We now show that this is still the case even under a positive curvature assumption. Namely, in Theorem 33, neither the assumption that $\sigma(x)^2$ is Lipschitz, nor the assumption that σ_∞ is bounded, can be removed (but see Remark 37).

Heavy tails with non-Lipschitz $\sigma(x)^2$. Our next example shows that if the diffusion constant $\sigma(x)^2$ is not Lipschitz, then non-exponential tails may occur in spite of positive curvature.

Consider the continuous-time random walk on $\mathbb{N}$ defined as follows: the transition rate from k to $k+1$ is $a(k+1)^2$, whereas the transition rate from k to $k-1$ is $a(k+1)^2 + bk$ for $k \geqslant 1$. Here $a, b > 0$ are fixed.

We have $\kappa = b$ and $\sigma(k)^2 = 2a(k+1)^2 + bk$, which is obviously not Lipschitz.

This Markov chain has a reversible measure ν, which satisfies $\nu(k)/\nu(k-1) = ak^2/(a(k+1)^2 + bk) = 1 - \frac{1}{k}(2 + \frac{b}{a}) + O(1/k^2)$. Consequently, asymptotically $\nu(k)$ behaves like

$$\prod_{i=1}^{k} \left(1 - \tfrac{1}{i}(2 + \tfrac{b}{a})\right) \approx \mathrm{e}^{-(2+b/a)\sum_{i=1}^{k} \frac{1}{i}} \approx k^{-(2+b/a)}$$

thus exhibiting heavy, non-exponential tails.

This shows that the Lipschitz assumption for $\sigma(x)^2$ cannot be removed, even though $\sigma_\infty = 1$. It would seem reasonable to expect a systematic correspondance between the asymptotic behavior of $\sigma(x)^2$ and the behavior of tails.

Heavy tails with unbounded σ_∞. Consider the following random walk on $\mathbb{N}^*$: a number k goes to 1 with probability $1 - 1/4k^2$ and to $2k$ with probability $1/4k^2$. One can check that $\kappa \geqslant 1/2$. These probabilities are chosen so that $\sigma(k)^2 = (2k-1)^2 \times 1/4k^2 \times (1 - 1/4k^2) \leqslant 1$, so that the variance of the invariant distribution is small. However, let us evaluate the probability that, starting at 1, the first i steps all consist in doing a multiplication by 2, so that we end at 2^i; this probability is $\prod_{j=0}^{i-1} \frac{1}{4\cdot(2^j)^2} = 4^{-1-i(i-1)/2}$. Setting $i = \log_2 k$, we see that the invariant distribution ν satisfies

$$\nu(k) \geqslant \frac{\nu(1)}{4}\, 2^{-\log_2 k\,(\log_2 k - 1)}$$

for k a power of 2. This is clearly not Gaussian or exponential, though $\sigma(k)^2$ is bounded.

4 Local control and logarithmic Sobolev inequality

We now turn to control of the gradient of $\mathrm{M}f$ at some point, in terms of the gradient of f at neighboring points. This is closer to classical Bakry–Émery theory, and allows to get a kind of logarithmic Sobolev inequality.

Definition 41 (Norm of the gradient).
Choose $\lambda > 0$ and, for any function $f : X \to \mathbb{R}$, define the λ-range gradient *of f by*

$$(\mathrm{D}f)(x) := \sup_{y,y' \in X} \frac{|f(y) - f(y')|}{d(y,y')}\, \mathrm{e}^{-\lambda d(x,y) - \lambda d(y,y')}$$

This is a kind of "mesoscopic" Lipschitz constant of f around x, since pairs of points y, y' far away from x will not contribute much to $\mathrm{D}f(x)$. If f is a smooth function on a compact Riemannian manifold, when $\lambda \to \infty$ this quantity tends to $|\nabla f(x)|$.

It is important to note that $\log \mathrm{D}f$ is λ-Lipschitz.

We will also need a control on negative curvature: In a Riemannian manifold, Ricci curvature might be $\geqslant 1$ because there is a direction of curvature 1000 and a direction of curvature -999. The next definition captures these variations.

Definition 42 (Unstability).
Let

$$\kappa_+(x,y) := \frac{1}{d(x,y)} \int_z (d(x,y) - d(x+z, y+z))_+$$

and

$$\kappa_-(x,y) := \frac{1}{d(x,y)} \int_z (d(x,y) - d(x+z, y+z))_-$$

where a_+ and a_- are the positive and negative part of $a \in \mathbb{R}$, so that $\kappa(x,y) = \kappa_+(x,y) - \kappa_-(x,y)$. (The integration over z is under a coupling realizing the value of $\kappa(x,y)$.)

The unstability $U(x,y)$ is defined as

$$U(x,y) := \frac{\kappa_-(x,y)}{\kappa(x,y)} \qquad \text{and} \qquad U := \sup_{x,y \in X,\, x \neq y} U(x,y)$$

Remark 43.
If X is ε-geodesic, then an upper bound for $U(x,y)$ with $d(x,y) \leqslant \varepsilon$ implies the same upper bound for U.

In most discrete examples given in the introduction (Examples 8, 10, 12, 13, 14), unstability is actually 0, meaning that the coupling between m_x and m_y never increases distances. (This could be a possible definition of non-negative *sectional* curvature for Markov chains.) In Riemannian manifolds, unstability is controlled by the largest negative sectional curvature. Interestingly, in Example 17 (Glauber dynamics), unstability depends on temperature.

Due to the use of the gradient D, the theorems below are interesting only if a reasonable estimate for $\mathrm{D}f$ can be obtained depending on “local” data. This is not the case when f is not λ-log-Lipschitz (compare the similar phenomenon in [BL98]). This is consistent with the fact mentioned above, that Gaussian concentration of measure only occurs in a finite range, with exponential concentration afterwards, which implies that no true logarithmic Sobolev inequality can hold in general.

Theorem 44 (Gradient contraction).
Suppose that coarse Ricci curvature is at least $\kappa > 0$. Let $\lambda \leqslant \frac{1}{20\sigma_\infty(1+U)}$ and consider the λ-range gradient D. Then for any function $f : X \to \mathbb{R}$ with $\mathrm{D}f < \infty$ we have

$$\mathrm{D}(\mathrm{M}f)(x) \leqslant (1 - \kappa/2)\mathrm{M}(\mathrm{D}f)(x)$$

for all $x \in X$.

Theorem 45 (Log-Sobolev inequality).
Suppose that coarse Ricci curvature is at least $\kappa > 0$. Let $\lambda \leqslant \frac{1}{20\sigma_\infty(1+U)}$ and consider the λ-range gradient D. Then for any function $f : x \to \mathbb{R}$ with $\mathrm{D}f < \infty$, we have

$$\mathrm{Var}_\nu f \leqslant \left(\sup_x \frac{4\sigma(x)^2}{\kappa n_x}\right) \int (\mathrm{D}f)^2 \,\mathrm{d}\nu$$

and for positive f,

$$\operatorname{Ent}_\nu f \leqslant \left(\sup_x \frac{4\sigma(x)^2}{\kappa n_x}\right) \int \frac{(\mathrm{D}f)^2}{f}\,\mathrm{d}\nu$$

where ν is the invariant distribution.

If moreover the random walk is reversible with respect to ν, then

$$\operatorname{Var}_\nu f \leqslant \int V(x)\,\mathrm{D}f(x)^2\,\mathrm{d}\nu(x)$$

and

$$\operatorname{Ent}_\nu f \leqslant \int V(x)\,\frac{\mathrm{D}f(x)^2}{f(x)}\,\mathrm{d}\nu(x)$$

where

$$V(x) = 2\sum_{t=0}^{\infty} (1-\kappa/2)^{2t} \mathrm{M}^{t+1}\!\left(\frac{\sigma(x)^2}{n_x}\right)$$

The form involving $V(x)$ is motivated by the fact that, for reversible diffusions in $\mathbb{R}^N$ with non-constant diffusion coefficients, the coefficients naturally appear in the formulation of functional inequalities (see e.g. [AMTU01]). The quantity $V(x)\,\mathrm{D}f(x)^2$ is to be thought of as a crude version of the Dirichlet form associated with the random walk. It would be more satisfying to obtain inequalities involving the latter (compare Corollary 31), but I could not get a version of the commutation property $\mathrm{DM} \leqslant (1-\kappa/2)\mathrm{MD}$ involving the Dirichlet form.

Remark 46.
If $\frac{\sigma(x)^2}{n_x\kappa}$ is C-Lipschitz (as in Theorem 33), then $V(x) \leqslant \frac{4\sigma^2}{\kappa n} + 2C\,\frac{J(x)}{\kappa}$.

Examples. Let us compare this theorem to classical results.

In the case of a Riemannian manifold, for any smooth function f we can choose a random walk with small enough steps, so that λ can be arbitrarily large and $\mathrm{D}f$ arbitrarily close to $|\nabla f|$. Since moreover $\sigma(x)^2$ does not depend on x for the Brownian motion, this theorem allows to recover the logarithmic Sobolev inequality in the Bakry–Émery framework, with the correct constant up to a factor 4.

Next, consider the two-point space $\{0,1\}$, equipped with the measure $\nu(0) = 1-p$ and $\nu(1) = p$. This is the space on which modified logarithmic Sobolev inequalities were introduced [BL98]. We endow this space with the Markov chain sending each point to the invariant distribution. Here we have $\sigma(x)^2 = p(1-p)$, $n_x = 1$ and $\kappa = 1$, so that we get the inequality $\operatorname{Ent}_\nu f \leqslant 4p(1-p)\int \frac{(\mathrm{D}f)^2}{f}\,\mathrm{d}\nu$, comparable to the known inequality [BL98] except for the factor 4.

The modified logarithmic Sobolev inequality for Bernoulli and Poisson measures is traditionally obtained by tensorizing this result [BL98]. If, instead, we directly apply the theorem above to the Bernoulli measure on $\{0,1\}^N$ or the Poisson measure on $\mathbb{N}$ (see Sections 3.3.3 and 3.3.4), we get slightly worse results. Indeed, consider the $M/M/\infty$ queue on $\mathbb{N}$, which is the limit when $N \to \infty$ of the projection on $\mathbb{N}$ of the

Markov chains on $\{0,1\}^N$ associated with Bernoulli measures. Keeping the notation of Section 3.3.4, we get, in the continuous-time version, $\sigma(x)^2 = x\mu + \lambda$, which is not bounded. So we have to use the version with $V(x)$; Remark 46 and the formulas in Section 3.3.4 yield $V(x) \leqslant 8\lambda/\mu + 2(\lambda + x\mu)/\mu$ so that we get the inequality

$$\begin{aligned}\mathrm{Ent}_\nu f &\leqslant \frac{\lambda}{\mu}\int \frac{\mathrm{D}f(x)^2}{f(x)}\,(10 + 2x\mu/\lambda)\,\mathrm{d}\nu(x)\\ &= \frac{\lambda}{\mu}\int \frac{\mathrm{D}f(x)^2}{f(x)}\,(2\,\mathrm{d}\nu(x-1) + 10\,\mathrm{d}\nu(x))\end{aligned}$$

which is to be compared to the inequality

$$\mathrm{Ent}_\nu f \leqslant \frac{\lambda}{\mu}\int \frac{\mathrm{D}_+f(x)^2}{f(x)}\,\mathrm{d}\nu(x)$$

obtained in [BL98], with $\mathrm{D}_+f(x) = f(x+1) - f(x)$. So asymptotically our version is worse by a factor $\mathrm{d}\nu(x-1)/\mathrm{d}\nu(x) \approx x$. One could say that our general, non-local notion of gradient fails to distinguish between a point and an immediate neighbor, and does not take advantage of the particular directional structure of a random walk on $\mathbb{N}$ as the use of D_+ does.

Yet being able to handle the configuration space directly rather than as a product of the two-point space allows us to deal with more general, non-product situations. Consider for example the queuing process with heterogeneous server rates mentioned at the end of Section 3.3.4, where newly arrived clients go to the fastest free server (in which case the number of busy servers is not Markovian). Then coarse Ricci curvature is equal to the infimum of the server rates, and Theorem 45 still holds, though the constants are probably not optimal when the rates are very different. I do not know if this result is new.

We now turn to the proof of Theorems 44 and 45. The proof of the former is specific to our setting, but the passage from the former to the latter is essentially a copy of the Bakry–Émery argument.

Lemma 47.
Let $x, y \in X$ with $\kappa(x,y) > 0$. Let (Z,μ) be a probability space equipped with a map $\pi : Z \to \operatorname{Supp} m_x \times \operatorname{Supp} m_y$ such that π sends μ to an optimal coupling between m_x and m_y. Let A be a positive function on Z such that $\sup A/\inf A \leqslant \mathrm{e}^\rho$ with $\rho \leqslant \frac{1}{2(1+U)}$. Then

$$\int_{z\in Z} A(z)\,\frac{d(x+z,y+z)}{d(x,y)} \leqslant (1-\kappa(x,y)/2)\int_z A(z)$$

and in particular

$$\int_{z\in Z} A(z)\,(d(x+z,y+z) - d(x,y)) \leqslant 0$$

where, as usual, $x+z$ and $y+z$ denote the two projections from Z to $\operatorname{Supp} m_x$ and $\operatorname{Supp} m_y$ respectively.

Proof.
The idea is the following: When A is constant, the result obviously holds since by definition $\int d(x+z,y+z)/d(x,y) = 1-\kappa(x,y)$. Now when A is close enough to a constant, the same holds with some numerical loss.

Set $F = \sup_z A(z)$. Then

$$\int_z A(z)\,\frac{d(x+z,y+z)}{d(x,y)} = \int_z A(z) + F\int_z \frac{A(z)}{F}\left(\frac{d(x+z,y+z)}{d(x,y)}-1\right)$$

Let $Z_- = \{z\in Z,\ d(x,y) < d(x+z,y+z)\}$ and $Z_+ = Z\backslash Z_-$. Recall that by definition, $\kappa_-(x,y) = \int_{Z_-}(d(x+z,y+z)/d(x,y)-1)$ and $\kappa_+(x,y) = \int_{Z_+}(1-d(x+z,y+z)/d(x,y))$, so that $\kappa = \kappa_+ - \kappa_-$. Using that $A(z)\leqslant F$ on Z_- and $A(z)\geqslant \mathrm{e}^{-\rho}F$ on Z_+, we get

$$\int_z A(z)\,\frac{d(x+z,y+z)}{d(x,y)} \leqslant \int_z A(z) + F(\kappa_-(x,y) - \mathrm{e}^{-\rho}\kappa_+(x,y))$$

Now by definition of U we have $\kappa_-(x,y)\leqslant U\kappa(x,y)$. It is not difficult to check that $\rho\leqslant\frac{1}{2(1+U)}$ is enough to ensure that $\mathrm{e}^{-\rho}\kappa_+(x,y)-\kappa_-(x,y)\geqslant\kappa(x,y)/2$, hence

$$\begin{aligned}\int_z A(z)\,\frac{d(x+z,y+z)}{d(x,y)} &\leqslant \int_z A(z) - F\kappa(x,y)/2\\ &\leqslant (1-\kappa(x,y)/2)\int_z A(z)\end{aligned}$$

as needed. □

Proof of Theorem 44.
Let $y,y'\in X$. Let ξ_{xy} and $\xi_{yy'}$ be optimal couplings between m_x and m_y, m_y and $m_{y'}$ respectively. Apply the gluing lemma for couplings (Lemma 7.6 in [Vil03]) to obtain a measure μ on $Z = \operatorname{Supp} m_x\times\operatorname{Supp} m_y\times\operatorname{Supp} m_{y'}$ whose projections on $\operatorname{Supp} m_x\times\operatorname{Supp} m_y$ and $\operatorname{Supp} m_y\times\operatorname{Supp} m_{y'}$ are ξ_{xy} and $\xi_{yy'}$ respectively.

We have

$$\begin{aligned}&\frac{|\mathrm{M}f(y)-\mathrm{M}f(y')|}{d(y,y')}\,\mathrm{e}^{-\lambda(d(x,y)+d(y,y'))}\\
&=\left|\int_{z\in Z} f(y+z)-f(y'+z)\right|\,\frac{\mathrm{e}^{-\lambda(d(x,y)+d(y,y'))}}{d(y,y')}\\
&\leqslant\int_{z\in Z}\left|f(y+z)-f(y'+z)\right|\,\frac{\mathrm{e}^{-\lambda(d(x,y)+d(y,y'))}}{d(y,y')}\\
&\leqslant\int_{z\in Z}\mathrm{D}f(x+z)\,\frac{d(y+z,y'+z)}{\mathrm{e}^{-\lambda(d(x+z,y+z)+d(y+z,y'+z))}}\,\frac{\mathrm{e}^{-\lambda(d(x,y)+d(y,y'))}}{d(y,y')}\\
&=\int_{z\in Z}A(z)B(z)\,\frac{d(y+z,y'+z)}{d(y,y')}\end{aligned}$$

where $A(z) = \mathrm{D}f(x+z)$ and $B(z) = \mathrm{e}^{\lambda(d(x+z,y+z)-d(x,y)+d(y+z,y'+z)-d(y,y'))}$.

Since $\operatorname{diam}\operatorname{Supp} m_x \leqslant 2\sigma_\infty$ and likewise for y, for any z, z' we have

$$\left|d(x+z,y+z)-d(x+z',y+z')\right| \leqslant 4\sigma_\infty$$
$$\left|d(y+z,y'+z)-d(y+z',y'+z')\right| \leqslant 4\sigma_\infty$$

so that B varies by a factor at most $\mathrm{e}^{8\lambda\sigma_\infty}$ on Z. Likewise, since $\mathrm{D}f$ is λ-log-Lipschitz, A varies by a factor at most $\mathrm{e}^{2\lambda\sigma_\infty}$. So the quantity $A(z)B(z)$ varies by at most $\mathrm{e}^{10\lambda\sigma_\infty}$.

So if $\lambda \leqslant \frac{1}{20\sigma_\infty(1+U)}$, we can apply Lemma 47 and get

$$\int_{z\in Z} A(z)B(z)\,\frac{d(y+z,y'+z)}{d(y,y')} \leqslant (1-\kappa/2)\int_{z\in Z} A(z)B(z)$$

Now we have $\int_z A(z)B(z) = \int_z A(z) + \int_z A(z)(B(z)-1)$. Unwinding $B(z)$ and using that $\mathrm{e}^a - 1 \leqslant a\mathrm{e}^a$ for any $a \in \mathbb{R}$, we get

$$\int_z A(z)(B(z)-1) \leqslant$$
$$\lambda\int_z A(z)B(z)\left(d(x+z,y+z)-d(x,y)+d(y+z,y'+z)-d(y,y')\right)$$

which decomposes as a sum of two terms $\lambda\int_z A(z)B(z)\,(d(x+z,y+z)-d(x,y))$ and $\lambda\int_z A(z)B(z)\,(d(y+z,y'+z)-d(y,y'))$, each of which is non-positive by Lemma 47. Hence $\int_z A(z)(B(z)-1) \leqslant 0$ and $\int_z A(z)B(z) \leqslant \int_z A(z) = \int_z \mathrm{D}f(x+z) = \mathrm{M}(\mathrm{D}f)(x)$. So we have shown that for any y, y' in X we have

$$\frac{|\mathrm{M}f(y)-\mathrm{M}f(y')|}{d(y,y')}\,\mathrm{e}^{-\lambda(d(x,y)+d(y,y'))} \leqslant (1-\kappa/2)\mathrm{M}(\mathrm{D}f)(x)$$

as needed. □

Lemma 48.
Let f be a function with $\mathrm{D}f < \infty$. Let $x \in X$. Then f is $\mathrm{e}^{4\lambda\sigma_\infty}\mathrm{M}(\mathrm{D}f)(x)$-Lipschitz on $\operatorname{Supp} m_x$.

Proof.
For any $y, z \in \operatorname{Supp} m_x$, by definition of D we have $|f(y)-f(z)| \leqslant \mathrm{D}f(y)d(y,z)\mathrm{e}^{\lambda d(y,z)} \leqslant \mathrm{D}f(y)d(y,z)\mathrm{e}^{2\lambda\sigma_\infty}$. Since moreover $\mathrm{D}f$ is λ-log-Lipschitz, we have $\mathrm{D}f(y) \leqslant \mathrm{e}^{2\lambda\sigma_\infty}\inf_{\operatorname{Supp} m_x}\mathrm{D}f \leqslant \mathrm{e}^{2\lambda\sigma_\infty}\mathrm{M}(\mathrm{D}f)(x)$, so that finally

$$|f(y)-f(z)| \leqslant d(y,z)\,\mathrm{M}(\mathrm{D}f)(x)\,\mathrm{e}^{4\lambda\sigma_\infty}$$

as announced. □

Proof of Theorem 45.
Let ν be the invariant distribution. Let f be a positive measurable function. Associativity of entropy (e.g. Theorem D.13 in [DZ98] applied to the measure $f(y)\,\mathrm{d}\nu(x)\mathrm{d}m_x(y)$

on $X \times X$) states that

$$\begin{aligned} \operatorname{Ent} f &= \int_x \operatorname{Ent}_{m_x} f \,\mathrm{d}\nu(x) + \operatorname{Ent} \mathrm{M} f \\ &= \sum_{t \geqslant 0} \int_x \operatorname{Ent}_{m_x} \mathrm{M}^t f \,\mathrm{d}\nu(x) \end{aligned}$$

by induction, and similarly

$$\operatorname{Var} f = \sum_{t \geqslant 0} \int_x \operatorname{Var}_{m_x} \mathrm{M}^t f \,\mathrm{d}\nu(x)$$

Since by the lemma above, f is $\mathrm{M}(\mathrm{D}f)(x)\,\mathrm{e}^{4\lambda\sigma_\infty}$-Lipschitz on $\operatorname{Supp} m_x$ and moreover $\mathrm{e}^{8\lambda\sigma_\infty} < 2$, we have

$$\operatorname{Var}_{m_x} f \leqslant \frac{2(\mathrm{M}(\mathrm{D}f)(x))^2\,\sigma(x)^2}{n_x}$$

and, using that $a \log a \leqslant a^2 - a$, we get that $\operatorname{Ent}_{m_x} f \leqslant \frac{1}{\mathrm{M}f(x)} \operatorname{Var}_{m_x} f$ so

$$\operatorname{Ent}_{m_x} f \leqslant \frac{2(\mathrm{M}(\mathrm{D}f)(x))^2\,\sigma(x)^2}{n_x\,\mathrm{M}f(x)}$$

Thus

$$\operatorname{Var} f \leqslant 2 \sum_{t \geqslant 0} \int_x \frac{\sigma(x)^2}{n_x} \,(\mathrm{M}(\mathrm{D}\mathrm{M}^t f)(x))^2 \,\mathrm{d}\nu(x)$$

and

$$\operatorname{Ent} f \leqslant 2 \sum_{t \geqslant 0} \int_x \frac{\sigma(x)^2}{n_x} \,\frac{(\mathrm{M}(\mathrm{D}\mathrm{M}^t f)(x))^2}{\mathrm{M}^{t+1} f(x)} \,\mathrm{d}\nu(x)$$

By Theorem 44, we have $(\mathrm{D}\mathrm{M}^t f)(y) \leqslant (1-\kappa/2)^t \mathrm{M}^t(\mathrm{D}f)(y)$, so that

$$\operatorname{Var} f \leqslant 2 \sum_{t \geqslant 0} \int_x \frac{\sigma(x)^2}{n_x} \,(\mathrm{M}^{t+1}\mathrm{D}f(x))^2 \,(1-\kappa/2)^{2t} \,\mathrm{d}\nu(x)$$

and

$$\operatorname{Ent} f \leqslant 2 \sum_{t \geqslant 0} \int_x \frac{\sigma(x)^2}{n_x} \,\frac{(\mathrm{M}^{t+1}\mathrm{D}f(x))^2}{\mathrm{M}^{t+1} f(x)} \,(1-\kappa/2)^{2t} \,\mathrm{d}\nu(x)$$

Now, for variance, convexity of $a \mapsto a^2$ yields

$$(\mathrm{M}^{t+1}\mathrm{D}f)^2 \leqslant \mathrm{M}^{t+1}((\mathrm{D}f)^2)$$

and for entropy, convexity of $(a, b) \mapsto a^2/b$ for $a, b > 0$ yields

$$\frac{(\mathrm{M}^{t+1}\mathrm{D}f(x))^2}{\mathrm{M}^{t+1} f(x)} \leqslant \mathrm{M}^{t+1}\left(\frac{(\mathrm{D}f)^2}{f}\right)(x)$$

Finally we get

$$\operatorname{Var} f \leqslant 2 \sum_{t \geqslant 0} (1-\kappa/2)^{2t} \int_x \frac{\sigma(x)^2}{n_x} \mathrm{M}^{t+1}((\mathrm{D}f)^2)(x) \, \mathrm{d}\nu(x)$$

and

$$\operatorname{Ent} f \leqslant 2 \sum_{t \geqslant 0} (1-\kappa/2)^{2t} \int_x \frac{\sigma(x)^2}{n_x} \mathrm{M}^{t+1} \left(\frac{(\mathrm{D}f)^2}{f} \right)(x) \, \mathrm{d}\nu(x)$$

Now, in the non-reversible case, simply apply the identity

$$\int g(x) \, \mathrm{M}^{t+1} h(x) \, \mathrm{d}\nu(x) \leqslant (\sup g) \int \mathrm{M}^{t+1} h(x) \, \mathrm{d}\nu(x) = (\sup g) \int h \, \mathrm{d}\nu$$

to the functions $g(x) = \frac{\sigma(x)^2}{n_x}$ and $h(x) = (\mathrm{D}f)(x)^2$ (for variance) or $h(x) = (\mathrm{D}f)(x)^2/f(x)$ (for entropy). For the reversible case, use the identity

$$\int g(x) \, \mathrm{M}^{t+1} h(x) \, \mathrm{d}\nu(x) = \int h(x) \, \mathrm{M}^{t+1} g(x) \, \mathrm{d}\nu(x)$$

instead. □

5 Exponential concentration in non-negative curvature

We have seen that positive coarse Ricci curvature implies a kind of Gaussian concentration. We now show that non-negative coarse Ricci curvature and the existence of an "attracting point" imply exponential concentration.

The basic example to keep in mind is the following. Let $\mathbb{N}$ be the set of non-negative integers equipped with its standard distance. Let $0 < p < 1$ and consider the nearest-neighbor random walk on $\mathbb{N}$ that goes to the left with probability p and to the right with probability $1-p$; explicitly $m_k = p\delta_{k-1} + (1-p)\delta_{k+1}$ for $k \geqslant 1$, and $m_0 = p\delta_0 + (1-p)\delta_1$.

Since for $k \geqslant 1$ the transition kernel is translation-invariant, it is immediate to check that $\kappa(k, k+1) = 0$; besides, $\kappa(0,1) = p$. There exists an invariant distribution if and only if $p > 1/2$, and it satisfies exponential concentration with decay distance $1/\log(p/(1-p))$. For $p = 1/2 + \varepsilon$ with small ε this behaves like $1/4\varepsilon$. Of course, when $p \leqslant 1/2$, there is no invariant distribution so that non-negative curvature alone does not imply concentration of measure.

Geometrically, what entails exponential concentration in this example is the fact that, for $p > 1/2$, the point 0 "pulls" its neighbor, and the pulling is transmitted by non-negative curvature. We now formalize this situation in the following theorem.

Theorem 49.
Let $(X, d, (m_x))$ be a locally compact metric space with random walk. Suppose that for some $o \in X$ and $r > 0$ one has:

- *$\kappa(x,y) \geqslant 0$ for all $x,y \in X$,*
- *for all $x \in X$ with $r \leqslant d(o,x) < 2r$, one has $W_1(m_x,\delta_o) < d(x,o)$,*
- *X is r-geodesic,*
- *there exists $s > 0$ such that each measure m_x satisfies the Gaussian-type Laplace transform inequality*

$$\mathbb{E}_{m_x} \mathrm{e}^{\lambda f} \leqslant \mathrm{e}^{\lambda^2 s^2/2} \mathrm{e}^{\lambda \mathbb{E}_{m_x} f}$$

  for any $\lambda > 0$ and any 1-Lipschitz function $f : \operatorname{Supp} m_x \to \mathbb{R}$.

Set $\rho = \inf\{d(x,o) - W_1(m_x,\delta_o),\ r \leqslant d(o,x) < 2r\}$ and assume $\rho > 0$.

Then there exists an invariant distribution for the random walk. Moreover, setting $D = s^2/\rho$ and $m = r + 2s^2/\rho + \rho(1 + J(o)^2/4s^2)$, for any invariant distribution ν we have

$$\int \mathrm{e}^{d(x,o)/D}\, \mathrm{d}\nu(x) \leqslant (4 + J(o)^2/s^2)\, \mathrm{e}^{m/D}$$

and so for any 1-Lipschitz function $f : X \to \mathbb{R}$ and $t \geqslant 0$ we have

$$\Pr\left(|f - f(o)| \geqslant t + m\right) \leqslant (8 + 2J(o)^2/s^2)\, \mathrm{e}^{-t/D}$$

So we get exponential concentration with characteristic decay distance s^2/ρ.

The last assumption is always satisfied with $s = 2\sigma_\infty$ (Proposition 1.16 in [Led01]).

Examples. Before proceeding to the proof, let us show how this applies to the geometric distribution above on $\mathbb{N}$. We take of course $o = 0$ and $r = 1$. We can take $s = 2\sigma_\infty = 2$. Now there is only one point x with $r \leqslant d(o,x) < 2r$, which is $x = 1$. It satisfies $m_1 = p\delta_0 + (1-p)\delta_2$, so that $W_1(m_1,\delta_0) = 2(1-p)$, which is smaller than $d(0,1) = 1$ if and only if $p > 1/2$ as was to be expected. So we can take $\rho = 1 - 2(1-p) = 2p - 1$. Then we get exponential concentration with characteristic distance $4/(2p-1)$. When p is very close to 1 this is not so good (because the discretization is too coarse), but when p is close to $1/2$ this is within a factor 2 of the optimal value.

Another example is the stochastic differential equation $\mathrm{d}X_t = S\,\mathrm{d}B_t - \alpha \frac{X_t}{|X_t|}\,\mathrm{d}t$ on $\mathbb{R}^n$, for which $\exp(-2\,|x|\,\alpha/S^2)$ is a reversible measure. Take as a Markov chain the Euler approximation scheme at time δt for this stochastic differential equation, as in Example 11. Taking $r = nS^2/\alpha$ yields that $\rho \geqslant \alpha\,\delta t/2$ after some simple computation. Since we have $s^2 = S^2\delta t$ for Gaussian measures at time δt, we get exponential concentration with decay distance $2S^2/\alpha$, which is correct up to a factor 4. The additive constant in the deviation inequality is $m = r + \rho(1 + J(o)^2/4s^2) + 2s^2/\rho$ which is equal to $(n+4)S^2/\alpha + O(\delta t)$ (note that $J(o)^2 \approx ns^2$). For comparison, the actual value for the average distance to the origin under the exponential distribution $\mathrm{e}^{-2|x|\alpha/S^2}$ is $nS^2/2\alpha$, so that up to a constant the dependency on dimension is recovered.

In general, the invariant distribution is not unique under the assumptions of the theorem. For example, start with the random walk on $\mathbb{N}$ above with geometric invariant distribution; now consider the disjoint union $\mathbb{N} \cup (\mathbb{N} + \frac{1}{2})$ where on $\mathbb{N} + \frac{1}{2}$ we use the same random walk translated by $\frac{1}{2}$: the assumptions are satisfied with $r = 1$ and $o = 0$, but clearly there are two disjoint invariant distributions. However, if $\kappa > 0$ in some large enough ball around o, then the invariant distribution will be unique.

Proof of the theorem.

Let us first prove a lemma which shows how non-negative curvature transmits the "pulling".

Lemma 50.
Let $x \in X$ with $d(x, o) \geqslant r$. Then $W_1(m_x, o) \leqslant d(x, o) - \rho$.

Proof.
If $d(o, x) < 2r$ then this is one of the assumptions. So we suppose that $d(o, x) \geqslant 2r$.

Since X is r-geodesic, let $o = y_0, y_1, y_2, \ldots, y_n = x$ be a sequence of points with $d(y_i, y_{i+1}) \leqslant r$ and $\sum d(y_i, y_{i+1}) = d(o, x)$. We can assume that $d(o, y_2) > r$ (otherwise, remove y_1). Set $z = y_1$ if $d(o, y_1) = r$ and $z = y_2$ if $d(o, y_1) < r$, so that $r \leqslant d(o, z) < 2r$. Now

$$\begin{aligned} W_1(\delta_o, m_x) &\leqslant W_1(\delta_o, m_z) + W_1(m_z, m_x) \\ &\leqslant d(o, z) - \rho + d(z, x) \end{aligned}$$

since $\kappa(z, x) \geqslant 0$. But $d(o, z) + d(z, x) = d(o, x)$ by construction, hence the conclusion. □

We are now ready to prove the theorem. The idea is to consider the function $\mathrm{e}^{\lambda d(x,o)}$. For points far away from the origin, since under the random walk the average distance to the origin decreases by ρ by the previous lemma, we expect the function to be multiplied by $\mathrm{e}^{-\lambda\rho}$ under the random walk operator. Close to the origin, the evolution of the function is controlled by the variance s^2 and the jump $J(o)$ of the origin. Since the integral of the function is preserved by the random walk operator, and it is multiplied by a quantity < 1 far away, this shows that the weight of faraway points cannot be too large.

More precisely, we need to tamper a little bit with what happens around the origin. Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be defined by $\varphi(x) = 0$ if $x < r$; $\varphi(x) = (x - r)^2/kr$ if $r \leqslant x < r(\frac{k}{2} + 1)$ and $\varphi(x) = x - r - kr/4$ if $x \geqslant r(\frac{k}{2} + 1)$, for some $k > 0$ to be chosen later. Note that φ is a 1-Lipschitz function and that $\varphi'' \leqslant 2/kr$.

If Y is any random variable with values in $\mathbb{R}_+$, we have

$$\mathbb{E}\varphi(Y) \leqslant \varphi(\mathbb{E}Y) + \frac{1}{2} \operatorname{Var} Y \sup \varphi'' \leqslant \varphi(\mathbb{E}Y) + \frac{1}{kr} \operatorname{Var} Y$$

Now choose some $\lambda > 0$ and consider the function $f : X \to \mathbb{R}$ defined by $f(x) = \mathrm{e}^{\lambda\varphi(d(o,x))}$. Note that $\varphi(d(o, x))$ is 1-Lipschitz, so that by the Laplace transform assumption we have

$$\mathrm{M}f(x) \leqslant \mathrm{e}^{\lambda^2 s^2/2} \mathrm{e}^{\lambda \mathrm{M}\varphi(d(o,x))}$$

The Laplace transform assumption implies that the variance under m_x of any 1-Lipschitz function is at most s^2. So by the remark above, we have

$$\mathrm{M}\varphi(d(o,x)) \leqslant \varphi(\mathrm{M}d(o,x)) + \frac{s^2}{kr} = \varphi(W_1(m_x,\delta_o)) + \frac{s^2}{kr}$$

so that finally

$$\mathrm{M}f(x) \leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr}\mathrm{e}^{\lambda\varphi(W_1(m_x,\delta_o))}$$

for any $x \in X$.

We will use different bounds on $\varphi(W_1(m_x,\delta_o))$ according to $d(o,x)$. First, if $d(x,o) < r$, then use non-negative curvature to write $W_1(m_x,\delta_o) \leqslant W_1(m_x,m_o) + J(o) \leqslant d(x,o) + J(o)$ so that $\varphi(W_1(m_x,\delta_o)) \leqslant \varphi(r+J(o)) \leqslant J(o)^2/kr$ so that

$$\mathrm{M}f(x) \leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr+\lambda J(o)^2/kr} = \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr+\lambda J(o)^2/kr} f(x)$$

since $f(x) = 1$.

Second, for any x with $d(x,o) \geqslant r$, Lemma 50 yields

$$\mathrm{M}f(x) \leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr}\mathrm{e}^{\lambda\varphi(d(x,o)-\rho)}$$

If $d(x,o) \geqslant r(\frac{k}{2}+1)+\rho$ then $\varphi(d(x,o)-\rho) = \varphi(d(x,o)) - \rho$ so that

$$\mathrm{M}f(x) \leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr-\lambda\rho} f(x)$$

If $r \leqslant d(x,o) < r(\frac{k}{2}+1)+\rho$, then $\varphi(d(x,o)-\rho) \leqslant \varphi(d(x,o))$ so that

$$\mathrm{M}f(x) \leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr} f(x)$$

Let ν be any probability measure such that $\int f\,\mathrm{d}\nu < \infty$. Let $X' = \{x \in X,\ d(x,o) < r(\frac{k}{2}+1)\}$ and $X'' = X \setminus X'$. Set $A(\nu) = \int_{X'} f\,\mathrm{d}\nu$ and $B(\nu) = \int_{X''} f\,\mathrm{d}\nu$. Combining the cases above, we have shown that

$$\begin{aligned}
&A(\nu * m) + B(\nu * m)\\
&\qquad = \int f\,\mathrm{d}(\nu * m) = \int \mathrm{M}f\,\mathrm{d}\nu\\
&\qquad = \int_{X'} \mathrm{M}f\,\mathrm{d}\nu + \int_{X''} \mathrm{M}f\,\mathrm{d}\nu\\
&\qquad \leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr+\lambda J(o)^2/kr}\int_{X'} f\,\mathrm{d}\nu + \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr-\lambda\rho}\int_{X''} f\,\mathrm{d}\nu\\
&\qquad = \alpha A(\nu) + \beta B(\nu)
\end{aligned}$$

with $\alpha = \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr+\lambda J(o)^2/kr}$ and $\beta = \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr-\lambda\rho}$.

Choose λ small enough and k large enough (see below) so that $\beta < 1$. Using that $A(\nu) \leqslant \mathrm{e}^{\lambda kr/4}$ for any probability measure ν, we get $\alpha A(\nu) + \beta B(\nu) \leqslant (\alpha-\beta)\mathrm{e}^{\lambda kr/4} + \beta(A(\nu)+B(\nu))$. In particular, if $A(\nu)+B(\nu) \leqslant \frac{(\alpha-\beta)\mathrm{e}^{\lambda kr/4}}{1-\beta}$, we get $\alpha A(\nu)+\beta B(\nu) \leqslant$

$\frac{(\alpha-\beta)e^{\lambda kr/4}}{1-\beta}$ again. So setting $R = \frac{(\alpha-\beta)e^{\lambda kr/4}}{1-\beta}$, we have just shown that the set C of probability measures ν such that $\int f\,\mathrm{d}\nu \leqslant R$ is invariant under the random walk.

Moreover, if $A(\nu)+B(\nu) > R$ then $\alpha A(\nu)+\beta B(\nu) < A(\nu)+B(\nu)$. Hence, if ν is an invariant distribution, necessarily $\nu \in C$. This, together with an evaluation of R given below, will provide the bound for $\int f\,\mathrm{d}\nu$ stated in the theorem.

We now turn to existence of an invariant distribution. First, C is obviously closed and convex. Moreover, C is tight: indeed if K is a compact containing a ball of radius a around o, then for any $\nu \in C$ we have $\nu(X \setminus K) \leqslant R\mathrm{e}^{-\lambda a}$. So by Prokhorov's theorem, C is compact in the weak convergence topology. So C is compact convex in the topological vector space of all (signed) Borel measures on X, and is invariant by the random walk operator, which is an affine map. By the Markov–Kakutani theorem (Theorem I.3.3.1 in [GD03]), it has a fixed point.

Let us finally evaluate R. We have

$$\begin{aligned}
R &= \frac{\alpha/\beta-1}{1/\beta-1}\,\mathrm{e}^{\lambda kr/4}\\
&= \frac{\mathrm{e}^{\lambda J(o)^2/kr+\lambda\rho}-1}{\mathrm{e}^{\lambda\rho-\lambda s^2/kr-\lambda^2 s^2/2}-1}\,\mathrm{e}^{\lambda kr/4}\\
&\leqslant \frac{\rho+J(o)^2/kr}{\rho-s^2/kr-\lambda s^2/2}\,\mathrm{e}^{\lambda J(o)^2/kr+\lambda\rho+\lambda kr/4}
\end{aligned}$$

using $\mathrm{e}^a-1 \leqslant a\mathrm{e}^a$ and $\mathrm{e}^a-1 \geqslant a$.

Now take $\lambda = \rho/s^2$ and $k = 4s^2/r\rho$. This yields

$$R \leqslant (4+J(o)^2/s^2)\,\mathrm{e}^{\lambda(s^2/\rho+\rho(1+J(o)^2/4s^2))}$$

Let ν be some invariant distribution: it satisfies $\int f\,\mathrm{d}\nu \leqslant R$. Since $d(x,o) \leqslant \varphi(d(x,o))+r(1+k/4)$ we have $\int \mathrm{e}^{\lambda d(x,o)}\,\mathrm{d}\nu \leqslant \mathrm{e}^{\lambda r(1+k/4)}\int f\,\mathrm{d}\nu \leqslant R\mathrm{e}^{\lambda r(1+k/4)}$, hence the result in the theorem. □

6 L^2 Bonnet–Myers theorems

As seen in Section 2.3, it is generally not possible to give a bound for the diameter of a positively curved space similar to the usual Bonnet–Myers theorem involving the square root of curvature, the simplest counterexample being the discrete cube. Here we describe additional conditions which provide such a bound in two different kinds of situation.

We first give a bound on the *average* distance between two points rather than the diameter; it holds when there is an "attractive point" and is relevant for examples such as the Ornstein–Uhlenbeck process (Example 9) or its discrete analogue (Example 10).

Next, we give a direct generalization of the genuine Bonnet–Myers theorem for Riemannian manifolds. Despite lack of further examples, we found it interesting to provide an axiomatization of the Bonnet–Myers theorem in our language. This is done by reinforcing the positive curvature assumption, which compares the transportation distance between the measures issuing from two points x and y at a given time, by requiring a transportation distance inequality between the measures issuing from two given points at *different* times.

6.1 Average L^2 Bonnet–Myers

We now describe a Bonnet–Myers–like estimate on the average distance between two points, provided there is some "attractive point". The proof is somewhat similar to that of Theorem 49.

Proposition 51 (Average L^2 Bonnet–Myers).
Let $(X, d, (m_x))$ be a metric space with random walk, with coarse Ricci curvature at least $\kappa > 0$. Suppose that for some $o \in X$ and $r \geqslant 0$, one has

$$W_1(\delta_o, m_x) \leqslant d(o, x)$$

for any $x \in X$ with $r \leqslant d(o, x) < 2r$, and that moreover X is r-geodesic.

Then

$$\int d(o, x)\, \mathrm{d}\nu(x) \leqslant \sqrt{\frac{1}{\kappa} \int \frac{\sigma(x)^2}{n_x}\, \mathrm{d}\nu(x)} + 5r$$

where as usual ν is the invariant distribution.

Note that the assumption $\int d(o, y)\, \mathrm{d}m_x(y) \leqslant d(o, x)$ cannot hold for x in some ball around o unless o is a fixed point. This is why the assumption is restricted to an annulus.

As in the Gaussian concentration theorem (Theorem 33), in case $\sigma(x)^2$ is Lipschitz, Corollary 22 may provide a useful bound on $\int \frac{\sigma(x)^2}{n_x}\, \mathrm{d}\nu(x)$ in terms of its value at some point.

As a first example, consider the discrete Ornstein–Uhlenbeck process of Example 10, which is the Markov chain on $\{-N, \ldots, N\}$ given by the transition probabilities $p_{k,k} = 1/2$, $p_{k,k+1} = 1/4 - k/4N$ and $p_{k,k-1} = 1/4 + k/4N$; the coarse Ricci curvature is $\kappa = 1/2N$, and the invariant distribution is the binomial $\frac{1}{2^{2N}}\binom{2N}{N+k}$. This example is interesting because the diameter is $2N$ (which is the bound provided by Proposition 23), whereas the average distance between two points is $\approx \sqrt{N}$. It is immediate to check that 0 is attractive, namely that $o = 0$ and $r = 1$ fulfill the assumptions. Since $\sigma(x)^2 \approx 1$ and $\kappa \approx 1/N$, the proposition recovers the correct order of magnitude for distance to the origin.

Our next example is the Ornstein–Uhlenbeck process $\mathrm{d}X_t = -\alpha\, X_t\, \mathrm{d}t + s\, \mathrm{d}B_t$ on $\mathbb{R}^N$ (Example 9). Here it is clear that 0 is attractive in some sense, so $o = 0$ is a natural choice. The invariant distribution is Gaussian of variance s^2/α; under this distribution the average distance to 0 is $\approx \sqrt{Ns^2/\alpha}$.

At small time τ, a point $x \in \mathbb{R}^N$ is sent to a Gaussian centered at $(1 - \alpha\tau)x$, of variance τs^2. The average quadratic distance to the origin under this Gaussian is $(1 - \alpha\tau)^2 d(0, x)^2 + Ns^2\tau + o(\tau)$ by a simple computation. If $d(0, x)^2 > Ns^2/2\alpha$ this is less than $d(0, x)^2$, so that we can take $r = \sqrt{Ns^2/2\alpha}$. Considering the random walk discretized at time τ we have we have $\kappa \sim \alpha\tau$, $\sigma(x)^2 \sim Ns^2\tau$ and $n_x \approx N$. So in the proposition above, the first term is $\approx \sqrt{s^2/\alpha}$, whereas the second term is $5r \approx \sqrt{Ns^2/\alpha}$, which is thus dominant. So the proposition gives the correct order of magnitude; in this precise case, the first term in the proposition reflects concentration of measure (which is dimension-independent for Gaussians), whereas it is the second term $5r$ which carries the correct dependency on dimension for the average distance to the origin.

Proof.

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be the function defined by $\varphi(x) = 0$ if $x \leqslant 2r$, and $\varphi(x) = (x - 2r)^2$ otherwise. For any real-valued random variable Y, we have

$$\mathbb{E}\varphi(Y) \leqslant \varphi(\mathbb{E}Y) + \frac{1}{2}\operatorname{Var} Y \sup \varphi'' = \varphi(\mathbb{E}Y) + \operatorname{Var} Y$$

Now let $f : X \to \mathbb{R}$ be defined by $f(x) = \varphi(d(o,x))$. We are going to show that

$$\mathrm{M}f(x) \leqslant (1-\kappa)^2 f(x) + \frac{\sigma(x)^2}{n_x} + 9r^2$$

for all $x \in X$. Since $\int f \,\mathrm{d}\nu = \int \mathrm{M}f \,\mathrm{d}\nu$, we will get $\int f \,\mathrm{d}\nu \leqslant (1-\kappa)^2 \int f \,\mathrm{d}\nu + \int \frac{\sigma(x)^2}{n_x} \,\mathrm{d}\nu + 9r^2$ which easily implies the result.

First, suppose that $r \leqslant d(o,x) < 2r$. We have $f(x) = 0$. Now $\int d(o,y) \,\mathrm{d}m_x(y)$ is at most $d(o,y)$ by assumption. Using the bound above for φ, together with the definition of $\sigma(x)^2$ and n_x, we get

$$\mathrm{M}f(x) = \int \varphi(d(o,y)) \,\mathrm{d}m_x(y) \leqslant \varphi\left(\int d(o,y) \,\mathrm{d}m_x(y)\right) + \frac{\sigma(x)^2}{n_x} = \frac{\sigma(x)^2}{n_x}$$

since $\int d(o,y) \,\mathrm{d}m_x(y) \leqslant 2r$ by assumption.

Second, suppose that $d(x,o) \geqslant 2r$. Using that X is r-geodesic, we can find a point x' such that $d(o,x) = d(o,x') + d(x',x)$ and $r \leqslant d(o,x') < 2r$ (take the second point in a sequence joining o to x). Now we have

$$\begin{aligned}
\int d(o,y) \,\mathrm{d}m_x(y) &= W_1(\delta_o, m_x) \\
&\leqslant W_1(\delta_o, m_{x'}) + W_1(m_{x'}, m_x) \\
&\leqslant W_1(\delta_o, m_{x'}) + (1-\kappa) d(x',x) \\
&\leqslant d(o,x') + (1-\kappa) d(x',x) \leqslant (1-\kappa) d(o,x) + 2\kappa r
\end{aligned}$$

and as above, this implies

$$\begin{aligned}
\mathrm{M}f(x) &\leqslant \varphi\left(\int d(o,y) \,\mathrm{d}m_x(y)\right) + \frac{\sigma(x)^2}{n_x} \\
&\leqslant ((1-\kappa) d(o,x) + 2\kappa r - 2r)^2 + \frac{\sigma(x)^2}{n_x} \\
&= (1-\kappa)^2 \varphi(d(o,x)) + \frac{\sigma(x)^2}{n_x}
\end{aligned}$$

as needed.

The last case to consider is $d(o,x) < r$. In this case we have

$$\begin{aligned}\int d(o,y)\,\mathrm{d}m_x(y) &= W_1(\delta_o, m_x)\\ &\leqslant W_1(\delta_o, m_o) + W_1(m_o, m_x) = J(o) + W_1(m_o, m_x)\\ &\leqslant J(o) + (1-\kappa)d(o,x) \leqslant J(o) + r\end{aligned}$$

So we need to bound $J(o)$. If X is included in the ball of radius r around o, the result trivially holds, so that we can assume that there exists a point x with $d(o,x) \geqslant r$. Since X is r-geodesic we can assume that $d(o,x) < 2r$ as well. Now $J(o) = W_1(m_o, \delta_o) \leqslant W_1(m_o, m_x) + W_1(m_x, \delta_o) \leqslant (1-\kappa)d(o,x) + W_1(m_x, \delta_o) \leqslant (1-\kappa)d(o,x) + d(o,x)$ by assumption, so that $J(o) \leqslant 4r$.

Plugging this into the above, for $d(o,x) < r$ we get $\int d(o,y)\,\mathrm{d}m_x(y) \leqslant 5r$ so that $\varphi(\int d(o,y)\,\mathrm{d}m_x(y)) \leqslant 9r^2$ hence $\mathrm{M}f(x) \leqslant 9r^2 + \frac{\sigma(x)^2}{n_x}$.

Combining the results, we get that whatever $x \in X$

$$\mathrm{M}f(x) \leqslant (1-\kappa)^2 f(x) + \frac{\sigma(x)^2}{n_x} + 9r^2$$

as needed. □

6.2 Strong L^2 Bonnet–Myers

As mentioned above, positive coarse Ricci curvature alone does not imply a $1/\sqrt{\kappa}$-like diameter control, because of such simple counterexamples as the discrete cube or the Ornstein–Uhlenbeck process. We now extract a property satisfied by the ordinary Brownian motion on Riemannian manifolds (without drift), which guarantees a genuine Bonnet–Myers theorem. Of course, this is of limited interest since the only available example is Riemannian manifolds, but nevertheless we found it interesting to find a sufficient condition expressed in our present language.

Our definition of coarse Ricci curvature controls the transportation distance between the measures issuing from two points x and x' at a given time t. The condition we will now use controls the transportation distance between the measures issuing from two points at two *different* times. It is based on what holds for Gaussian measures in $\mathbb{R}^N$. For any $x, x' \in \mathbb{R}^N$ and $t, t' > 0$, let m_x^{*t} and $m_{x'}^{*t'}$ be the laws of the standard Brownian motion issuing from x at time t and from x' at time t', respectively. It is easy to check that the L^2 transportation distance between these two measures is

$$W_2(m_x^{*t}, m_{x'}^{*t'})^2 = d(x,x')^2 + N(\sqrt{t} - \sqrt{t'})^2$$

hence

$$W_1(m_x^{*t}, m_{x'}^{*t'}) \leqslant d(x,x') + \frac{N(\sqrt{t} - \sqrt{t'})^2}{2d(x,x')}$$

The important feature here is that, when t' tends to t, the second term is of second order in $t' - t$. This is no more the case if we add a drift term to the diffusion.

We now take this inequality as an assumption and use it to copy the traditional proof of the Bonnet–Myers theorem. Here, for simplicity of notation we suppose that we are given a continuous-time Markov chain; however, the proof uses only a finite number of different values of t, so that discretization is possible (this is important in Riemannian manifolds, because the heat kernel is positive on the whole manifold at any positive time, and there is no simple control on it far away from the initial point; taking a discrete approximation with bounded steps solves this problem).

Proposition 52 (Strong L^2 Bonnet–Myers).
Let X be a metric space equipped with a continuous-time random walk m^{*t}. Assume that X is ε-geodesic, and that there exist constants $\kappa > 0, C \geqslant 0$ such that for any two small enough t, t', for any $x, x' \in X$ with $\varepsilon \leqslant d(x,x') \leqslant 2\varepsilon$ one has

$$W_1(m_x^{*t}, m_{x'}^{*t'}) \leqslant \mathrm{e}^{-\kappa \inf(t,t')} d(x,x') + \frac{C(\sqrt{t}-\sqrt{t'})^2}{2d(x,x')}$$

with $\kappa > 0$. Assume moreover that $\varepsilon \leqslant \frac{1}{2}\sqrt{C/2\kappa}$.

Then

$$\operatorname{diam} X \leqslant \pi \sqrt{\frac{C}{2\kappa}} \left(1 + \frac{4\varepsilon}{\sqrt{C/2\kappa}}\right)$$

When $t = t'$, the assumption reduces to $W_1(m_x^{*t}, m_{x'}^{*t}) \leqslant \mathrm{e}^{-\kappa t} d(x,x')$, which is just the continuous-time version of the positive curvature assumption. The constant C plays the role of a diffusion constant, and is equal to N for (a discrete approximation of) Brownian motion on a Riemannian manifold. We restrict the assumption to $d(x,x') \geqslant \varepsilon$ to avoid divergence problems for $\frac{C(\sqrt{t}-\sqrt{t'})^2}{2d(x,x')}$ when $x' \to x$.

For Brownian motion on an N-dimensional Riemannian manifold, we can take $\kappa = \frac{1}{2} \inf \mathrm{Ric}$ by Bakry-Émery theory (the $\frac{1}{2}$ is due to the fact that the infinitesimal generator of Brownian motion is $\frac{1}{2}\Delta$), and $C = N$ as in $\mathbb{R}^N$. So we get the usual Bonnet–Myers theorem, up to a factor $\sqrt{N}$ instead of $\sqrt{N-1}$ (similarly to our spectral gap estimate in comparison with the Lichnerowicz theorem), but with the correct constant π.

Proof.
Let $x, x' \in X$. Since X is ε-geodesic, we can find a sequence $x = x_0, x_1, \ldots, x_{k-1}, x_k = x'$ of points in X with $d(x_i, x_{i+1}) \leqslant \varepsilon$ and $\sum d(x_i, x_{i+1}) = d(x_0, x_k)$. By taking a subsequence (denoted x_i again), we can assume that $\varepsilon \leqslant d(x_i, x_{i+1}) \leqslant 2\varepsilon$ instead.

Set $t_i = \eta \sin\left(\frac{\pi d(x,x_i)}{d(x,x')}\right)^2$ for some (small) value of η to be chosen later. Now, since $t_0 = t_k = 0$ we have

$$\begin{aligned} d(x,x') = W_1(\delta_x, \delta_{x'}) &\leqslant \sum W_1(m_{x_i}^{*t_i}, m_{x_{i+1}}^{*t_{i+1}}) \\ &\leqslant \sum \mathrm{e}^{-\kappa \inf(t_i, t_{i+1})} d(x_i, x_{i+1}) + \frac{C(\sqrt{t_{i+1}}-\sqrt{t_i})^2}{2d(x_i, x_{i+1})} \end{aligned}$$

by assumption. Now we have $|\sin b - \sin a| = \left|2\sin\frac{b-a}{2}\cos\frac{a+b}{2}\right| \leqslant |b-a|\left|\cos\frac{a+b}{2}\right|$ so that

$$\frac{C(\sqrt{t_{i+1}}-\sqrt{t_i})^2}{2d(x_i,x_{i+1})} \leqslant \frac{C\eta\pi^2 d(x_i,x_{i+1})}{2d(x,x')^2}\cos^2\left(\pi\frac{d(x,x_i)+d(x,x_{i+1})}{2d(x,x')}\right)$$

Besides, if η is small enough, one has $\mathrm{e}^{-\kappa\inf(t_i,t_{i+1})} = 1-\kappa\inf(t_i,t_{i+1})+O(\eta^2)$. So we get

$$\begin{aligned} d(x,x') \leqslant \sum d(x_i,x_{i+1}) &- \kappa\inf(t_i,t_{i+1})d(x_i,x_{i+1}) \\ &+ \frac{C\eta\pi^2 d(x_i,x_{i+1})}{2d(x,x')^2}\cos^2\left(\pi\frac{d(x,x_i)+d(x,x_{i+1})}{2d(x,x')}\right) + O(\eta^2) \end{aligned}$$

Now the terms $\sum d(x_i,x_{i+1})\cos^2\left(\pi\frac{d(x,x_i)+d(x,x_{i+1})}{2d(x,x')}\right)$ and $\sum \inf(t_i,t_{i+1})d(x_i,x_{i+1})$ are close to the integrals $d(x,x')\int_0^1\cos^2(\pi u)\,\mathrm{d}u$ and $d(x,x')\eta\int_0^1\sin^2(\pi u)\,\mathrm{d}u$ respectively; the relative error in the Riemann sum is easily bounded by $\pi\varepsilon/d(x,x')$ so that

$$\begin{aligned} d(x,x') \leqslant d(x,x') &- \kappa\,\eta\,d(x,x')\left(\frac{1}{2}-\frac{\pi\varepsilon}{d(x,x')}\right) \\ &+ \frac{C\eta\pi^2}{2d(x,x')^2}\,d(x,x')\left(\frac{1}{2}+\frac{\pi\varepsilon}{d(x,x')}\right) + O(\eta^2) \end{aligned}$$

hence, taking η small enough,

$$d(x,x')^2 \leqslant \frac{C\pi^2}{2\kappa}\,\frac{1+2\pi\varepsilon/d(x,x')}{1-2\pi\varepsilon/d(x,x')}$$

so that either $d(x,x') \leqslant \pi\sqrt{C/2\kappa}$, or $2\pi\varepsilon/d(x,x') \leqslant 2\pi\varepsilon/\pi\sqrt{C/2\kappa} \leqslant 1/2$ by the assumption that ε is small, in which case we use $(1+a)/(1-a) \leqslant 1+4a$ for $a \leqslant 1/2$, hence the conclusion. □

7 Coarse Ricci curvature and Gromov–Hausdorff topology

One of our goals was to define a *robust* notion of curvature, not relying on differential calculus or the small-scale structure of a space. Here we first give two remarks about how changes to the metric and the random walk affect curvature. Next, in order to be able to change the underlying space as well, we introduce a Gromov–Hausdorff–like topology for metric spaces equipped with a random walk.

First, since coarse Ricci curvature is defined as a ratio between a transportation distance and a distance, we get the following remark.

Remark 53 (Change of metric).
Let $(X,d,m=(m_x))$ be a metric space with random walk, and let d' be a metric on X which is bi-Lipschitz equivalent to d, with constant $C \geqslant 1$. Suppose that the coarse

Ricci curvature of m on (X,d) is at least κ. Then the coarse Ricci curvature of m on (X,d') is at least κ' where $1-\kappa' = C^2(1-\kappa)$.

As an example, consider the ε-step random walk on a Riemannian manifold with positive Ricci curvature; κ behaves like ε^2 times the usual Ricci curvature, so that small bi-Lipschitz deformations of the metric, smaller than $O(\varepsilon^2)$, will preserve positivity of curvature of the ε-step random walk.

The next remark states that we can deform the random walk $m = (m_x)$ if the deformation depends on x in a Lipschitz way. Given a metric space (X,d), consider the space of 0-mass signed measures $\mathcal{P}_0(X) = \{\mu_+ - \mu_-\}$ where μ_+, μ_- are measures on X with finite first moment and with the same total mass. Equip this space with the norm (it is one) $\|\mu_+ - \mu_-\| := \sup_{f \text{ 1-Lipschitz}} \int f \,\mathrm{d}(\mu_+ - \mu_-) = W_1(\mu_+, \mu_-)$. Then the following trivially holds.

Remark 54 (Change of random walk).
Let (X,d) be a metric space and let $m = (m_x)_{x\in X}$, $m' = (m'_x)_{x\in X}$ be two random walks on X. Suppose that the coarse Ricci curvature of m is at least κ, and that the map $x \mapsto m_x - m'_x \in \mathcal{P}_0(X)$ is C-Lipschitz. Then the coarse Ricci curvature of m' is at least $\kappa - 2C$.

We now turn to changes in the space itself, for which we need to give a generalization of Gromov–Hausdorff topology taking the random walk data into account. Two spaces are close in this topology if they are close in the Gromov–Hausdorff topology and if moreover, the measures issuing from each point x are (uniformly) close in the L^1 transportation distance.

Recall [BBI01] that two metric spaces (X, d_X) and (Y, d_Y) are at Gromov–Hausdorff distance at most $e \in [0;\infty]$ if there exists a semi-metric space (Z, d_Z) and isometries $f_X : X \hookrightarrow Z$, $f_Y : Y \hookrightarrow Z$, such that for any $x \in X$, there exists $y \in Y$ with $d_Z(f_X(x), f_Y(y)) \leqslant e$, and likewise for any $y \in Y$ (i.e. the Hausdorff distance between $f_X(X)$ and $f_Y(Y)$ is at most e). We extend this definition as follows to incorporate the random walk.

Definition 55.
Let $(X, (m_x)_{x\in X})$ and $(Y, (m_y)_{y\in Y})$ be two metric spaces equipped with a random walk. For $e \in [0;\infty]$, we say that these spaces are e-close if there exists a metric space Z and two isometric embeddings $f_X : X \hookrightarrow Z$, $f_Y : Y \hookrightarrow Z$ such that for any $x \in X$, there exists $y \in Y$ such that $d_Z(f_X(x), f_Y(y)) \leqslant e$ and the L^1 transportation distance between the pushforward measures $f_X(m_x)$ and $f_Y(m_y)$ is at most $2e$, and likewise for any $y \in Y$.

It is easy to see that this is defines a semi-metric on the class of metric spaces equipped with a random walk. We say that a sequence of spaces with random walks $\big(X^N, (m^N_x)_{x\in X^N}\big)$ *converges* to $(X, (m_x))$ if the semi-distance between $(X^N, (m^N_x))$ and (X, m_x) tends to 0. We say, moreover, that a sequence of points $x^N \in X^N$ tends to $x \in X$ if we can take x^N and x to be corresponding points in the definition above. We give a similar definition for convergence of tuples of points in X^N.

Coarse Ricci curvature is a continuous function in this topology. Namely, a limit of spaces with coarse Ricci curvature at least κ has coarse Ricci curvature at least κ, as expressed in the following proposition.

Proposition 56 (Gromov–Hausdorff continuity).
Let $\left(X^N, (m_x^N)_{x\in X^N}\right)$ be a sequence of metric spaces with random walk, converging to a metric space with random walk $(X, (m_x)_{x\in X})$. Let x, y be two distinct points in X and let $(x^N, y^N) \in X^N \times X^N$ be a sequence of pairs of points converging to (x, y). Then $\kappa(x^N, y^N) \to \kappa(x, y)$.

In particular, if all spaces X^N have coarse Ricci curvature at least κ, then so does X. Thus, having coarse Ricci curvature at least κ is a closed property.

Proof.
We have $\kappa(x, y) = 1 - \frac{W_1(m_x,m_y)}{d(x,y)}$ and likewise for $\kappa(x^N, y^N)$. The definition ensures that $d(x^N, y^N)$ and $W_1(m_x^N, m_y^N)$ tend to $d(x, y)$ and $W_1(m_x, m_y)$ respectively, hence the result. □

Note however, that the coarse Ricci curvature of $(X, (m_x))$ may be larger than the limsup of the coarse Ricci curvatures of $(X^N, (m_x^N))$, because pairs of points in X^N, contributing to the curvature of X^N, may tend to the same point in X; for example, X may consist of a single point.

This collapsing phenomenon prevents positive curvature from being an *open* property. Yet it is possible to relax the definition of coarse Ricci curvature so as to allow any variation at small scales; with this perturbed definition, having coarse Ricci curvature greater than κ will become an open property. This is achieved as follows (compare the passage from trees to δ-hyperbolic spaces).

Definition 57.
Let (X, d) be a metric space equipped with a random walk m. Let $\delta \geqslant 0$. The coarse Ricci curvature up to δ along $x, y \in X$ *is the largest $\kappa \leqslant 1$ for which*

$$W_1(m_x, m_y) \leqslant (1 - \kappa)d(x, y) + \delta$$

With this definition, the following is easy.

Proposition 58.
Let $(X, (m_x))$ be a metric space with random walk with coarse Ricci curvature at least κ up to $\delta \geqslant 0$. Let $\delta' > 0$. Then there exists a neighborhood $\mathcal{V}_X$ of X such that any space $Y \in \mathcal{V}_X$ has coarse Ricci curvature at least κ up to $\delta + \delta'$.

Consequently, the property "having curvature at least κ for some $\delta \in [0; \delta_0)$" is open.

It would be interesting to study which properties of positive coarse Ricci curvature carry to this more general setting.

8 Transportation distance in Riemannian manifolds

Here we give the proofs of Proposition 6 and of the statements of Example 7 and Section 3.3.1.

We begin with Proposition 6 and evaluation of the coarse Ricci curvature of the ε-step random walk. The argument is close to the one in [RS05] (Theorem 1.5 (*xii*)), except that we use the value of Ricci curvature at a given point instead of its infimum on the manifold.

Let X be a smooth N-dimensional Riemannian manifold and let $x \in X$. Let v, w be unit tangent vectors at x. Let $\delta, \varepsilon > 0$ small enough. Let $y = \exp_x(\delta v)$. Let $x' = \exp_x(\varepsilon w)$ and $y' = \exp_y(\varepsilon w')$ where w' is the tangent vector at y obtained by parallel transport of w along the geodesic $t \mapsto \exp_x(tv)$. The first claim is that $d(x', y') = \delta\left(1 - \frac{\varepsilon^2}{2} K(v, w) + O(\delta\varepsilon^2 + \varepsilon^3)\right)$.

We suppose for simplicity that w and w' are orthogonal to v.

We will work in cylindrical coordinates along the geodesic $t \mapsto \exp_x(tv)$. Let $v_t = \frac{\mathrm{d}}{\mathrm{d}t} \exp_x(tv)$ be the speed of this geodesic. Let E_t be the orthogonal of v_t in the tangent space at $\exp_x(tv)$. Each point z in some neighborhood of x can be uniquely written as $\exp_{\exp_x(\tau(z)v)}(\varepsilon\zeta(z))$ for some $\tau(z) \in \mathbb{R}$ and $\zeta(z) \in E_{\tau(z)}$.

Consider the set $\exp_x(E_0)$ (restricted to some neighborhood of x to avoid topological problems), which contains x'. Let γ be a geodesic starting at some point of $\exp_x(E_0)$ and ending at y', which realizes the distance from $\exp_x(E_0)$ to y'. The distance from x' to y' is at least the length of γ. If δ and ε are small enough, the geodesic γ is arbitrarily close to the Euclidean one so that the coordinate τ is strictly increasing along γ. Let us parametrize γ using the coordinate τ, so that $\tau(\gamma(t)) = t$. Let also $w_t = \zeta(\gamma(t)) \in E_t$. In particular $w_\delta = w'$.

Consider, for each t, the geodesic $c_t : s \mapsto \exp_{\exp_x(tv)}(sw_t)$. We have $\gamma(t) = c_t(\varepsilon)$. For each given t, the vector field $\frac{D}{dt} c_t(s)$ is a Jacobi field along the geodesic $s \mapsto c_t(s)$. The initial conditions of this Jacobi field for $s = 0$ are given by $\frac{D}{\mathrm{d}t} c_t(s)_{|s=0} = v_t$ and $\frac{D}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}s} c_t(s)_{|s=0} = \frac{D}{\mathrm{d}t} w_t$. Applying the Jacobi equation yields

$$\left|\frac{\mathrm{d}\gamma(t)}{\mathrm{d}t}\right|^2 = \left|\frac{\mathrm{d}c_t(\varepsilon)}{\mathrm{d}t}\right|^2 = |v_t|^2 + 2\varepsilon\langle v_t, \dot{w}_t\rangle + \varepsilon^2 |\dot{w}_t|^2 - \varepsilon^2 \langle R(w_t, v_t)w_t, v_t\rangle + O(\varepsilon^3)$$

where $\dot{w}_t = \frac{D}{\mathrm{d}t} w_t$. But since by definition $w_t \in E_t$, we have $\langle v_t, \dot{w}_t\rangle = 0$. Since moreover $|v_t| = 1$ we get

$$\left|\frac{\mathrm{d}\gamma(t)}{\mathrm{d}t}\right| = 1 + \frac{\varepsilon^2}{2} |\dot{w}_t|^2 - \frac{\varepsilon^2}{2} \langle R(w_t, v_t)w_t, v_t\rangle + O(\varepsilon^3)$$

Integrating from $t = 0$ to $t = \delta$ and using that $\langle R(w_t, v_t)w_t, v_t\rangle = K(w, v) + O(\delta)$ yields that the length of the geodesic γ is

$$\delta\left(1 - \frac{\varepsilon^2}{2} K(v, w) + O(\varepsilon^3) + O(\varepsilon^2\delta)\right) + \frac{\varepsilon^2}{2}\int_{t=0}^{\delta} |\dot{w}_t|^2$$

so that the minimal value is achieved for $\dot{w}_t = 0$. But by definition $\dot{w}_t = 0$ means that the geodesic γ starts at x'. So first, we have estimated $d(x', y')$, which proves Proposition 6, and second, we have proven that the distance from y' to $\exp_x(E_0)$ is realized by x' up to the higher-order terms, which we will use below.

Let us now prove the statement of Example 7. Let μ_0, μ_1 be the uniform probability measures on the balls of radius ε centered at x and y respectively. We have to prove that

$$W_1(\mu_0, \mu_1) = d(x, y)\left(1 - \frac{\varepsilon^2}{2(N+2)}\,\mathrm{Ric}(v, v)\right)$$

up to higher-order terms.

Let μ_0', μ_1' be the images under the exponential map, of the uniform probability measures on the balls of radius ε in the tangent spaces at x and y respectively. So μ_0' is a measure having density $1 + O(\varepsilon^2)$ w.r.t. μ_0, and likewise for μ_1'.

If we average Proposition 6 over w in the ball of radius ε in the tangent space at x, we get that

$$W_1(\mu_0', \mu_1') \leqslant d(x, y)\left(1 - \frac{\varepsilon^2}{2(N+2)}\,\mathrm{Ric}(v, v)\right)$$

up to higher-order terms, since the coupling by parallel transport realizes this value. Indeed, the average of $K(v, w)$ on the unit sphere of the tangent plane at x is $\frac{1}{N}\,\mathrm{Ric}(v, v)$. Averaging on the ball instead of the sphere yields an $\frac{1}{N+2}$ factor instead.

Now the density of μ_0', μ_1' with respect to μ_0, μ_1 is $1 + O(\varepsilon^2)$. More precisely write $\frac{\mathrm{d}\mu_0'}{\mathrm{d}\mu_0} = 1 + \varepsilon^2 f_0$ and $\frac{\mathrm{d}\mu_1'}{\mathrm{d}\mu_1} = 1 + \varepsilon^2 f_1$ (where f_0 and f_1 can be written very explicitly in terms of the metric and its derivatives). Note that $f_1 = f_0 + O(d(x, y))$, and that moreover f_0 integrates to 0 since both μ_0 and μ_0' are probability measures. Plugging all this in the estimate above, we get the inequality for $W_1(\mu_0, \mu_1)$ up to the desired higher-order terms.

The converse inequality is proven as follows: if f is any 1-Lipschitz function, the L^1 transportation distance between measures μ_0 and μ_1 is at least the difference of the integrals of f under μ_0 and μ_1. Consider the function f equal to the distance of a point to $\exp_x(E_0)$ (taken in some small enough neighborhood of x), equipped with a $-$ sign if the point is not on the same side of E_0 as y. Clearly f is 1-Lipschitz. We computed above a lower bound for f in cylindrical coordinates, which after integrating yields a lower bound for $W_1(\mu_0', \mu_1')$. Arguments similar to the above turns this into the desired lower bound for $W_1(\mu_0, \mu_1)$.

Finally, let us briefly sketch the proofs of the other statements of Section 3.3.1, namely, evaluation of the diffusion constant and local dimension (Definition 18). Up to a multiplicative factor $1 + O(\varepsilon)$, these can be computed in the Euclidean space.

A simple computation shows that the expectation of the square distance of two points taken at random in a ball of radius ε in $\mathbb{R}^N$ is $\varepsilon^2\,\frac{2N}{N+2}$, hence the value $\varepsilon^2\,\frac{N}{N+2}$ for the diffusion constant $\sigma(x)^2$.

To evaluate the local dimension $n_x = \frac{\sigma(x)^2}{\sup \mathrm{Var}_{m_x} f,\ f \text{ 1-Lipschitz}}$ (Definition 18), we have to bound the maximal variance of a 1-Lipschitz function on a ball of radius ε in

$\mathbb{R}^N$. We will prove that the local dimension n_x is comprised between $N-1$ and N. A projection to a coordinate axis provides a function with variance $\frac{\varepsilon^2}{N+2}$, so that local dimension is at most N. For the other bound, let f be a 1-Lipschitz function on the ball and let us compute an upper bound for its variance. Take $\varepsilon = 1$ for simplicity. Write the ball of radius 1 as the union of the spheres S_r of radii $r \leqslant 1$. Let $v(r)$ be the variance of f restricted to the sphere S_r, and let $a(r)$ be the average of f on S_r. Then associativity of variances gives

$$\operatorname{Var} f = \int_{r=0}^{1} v(r)\,\mathrm{d}\mu(r) + \operatorname{Var}_\mu a(r)$$

where μ is the measure on the interval $[0;1]$ given by $\frac{r^{N-1}}{Z}\,\mathrm{d}r$ with $Z = \int_{r=0}^{1} r^{N-1}\,\mathrm{d}r = \frac{1}{N}$.

Since the variance of a 1-Lipschitz function on the $(N-1)$-dimensional unit sphere is at most $\frac{1}{N}$, we have $v(r) \leqslant \frac{r^2}{N}$ so that $\int_{r=0}^{1} v(r)\,\mathrm{d}\mu(r) \leqslant \frac{1}{N+2}$. To evaluate the second term, note that $a(r)$ is again 1-Lipschitz as a function of r, so that $\operatorname{Var}_\mu a(r) = \frac{1}{2}\iint (a(r)-a(r'))^2\,\mathrm{d}\mu(r)\mathrm{d}\mu(r')$ is at most $\frac{1}{2}\iint (r-r')^2\,\mathrm{d}\mu(r)\mathrm{d}\mu(r') = \frac{N}{(N+1)^2(N+2)}$. So finally

$$\operatorname{Var} f \leqslant \frac{1}{N+2} + \frac{N}{(N+1)^2(N+2)}$$

so that the local dimension n_x is bounded below by $\frac{N(N+1)^2}{N^2+3N+1} \geqslant N-1$.

References

[ABCFGMRS00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses **10**, Société Mathématique de France (2000).

[AMTU01] A. Arnold, P. Markowich, G. Toscani, A. Unterreiter, *On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations*, Comm. Partial Differential Equations **26** (2001), n° 1–2, 43–100.

[BBI01] D. Burago, Y. Burago, S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics **33**, AMS (2001).

[BD97] R. Bubley, M. E. Dyer, *Path coupling: a technique for proving rapid mixing in Markov chains*, FOCS 1997, 223–231.

[BE84] D. Bakry, M. Émery, *Hypercontractivité de semi-groupes de diffusion*, C. R. Acad. Sci. Paris Sér. I Math. **299** (1984), n° 15, 775–778.

[BE85] D. Bakry, M. Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84. Lecture Notes in Math. **1123**, Springer, Berlin (1985), 177–206.

[Ber03] M. Berger, *A panoramic view of Riemannian geometry*, Springer, Berlin (2003).

[BL98] S. Bobkov, M. Ledoux, *On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures*, J. Funct. Anal. **156** (1998), n° 2, 347–365.

[Bré99] P. Brémaud, *Markov chains*, Texts in Applied Mathematics **31**, Springer, New York (1999).

[BS] A.-I. Bonciocat, K.-T. Sturm, *Mass transportation and rough curvature bounds for discrete spaces*, preprint.

[Che98] M.-F. Chen, *Trilogy of couplings and general formulas for lower bound of spectral gap*, in *Probability towards 2000 (New York, 1995)*, Lecture Notes in Statist. **128**, Springer, New York (1998), 123–136.

[Che04] M.-F. Chen, *From Markov chains to non-equilibrium particle systems*, second edition, World Scientific (2004).

[CL89] M.-F. Chen, S.-F. Li, *Coupling methods for multidimensional diffusion processes*, Ann. Probab. **17** (1989), n° 1, 151–177.

[CW94] M.-F. Chen, F.-Y. Wang, *Application of coupling method to the first eigenvalue on manifold*, Sci. China Ser. A **37** (1994), n° 1, 1–14.

[DGW04] H. Djellout, A. Guillin, L. Wu, *Transportation cost-information inequalities and applications to random dynamical systems and diffusions*, Ann. Prob. **32** (2004), n° 3B, 2702–2732.

[Dob56a] R. L. Dobrušin, *On the condition of the central limit theorem for inhomogeneous Markov chains* (Russian), Dokl. Akad. Nauk SSSR (N.S.) **108** (1956), 1004–1006.

[Dob56b] R. Dobrushin, *Central limit theorem for non-stationary Markov chains I* (Russian), Teor. Veroyatnost. i Primenen. **1** (1956), 72–89.

[Dob70] R. L. Dobrušin, *Definition of a system of random variables by means of conditional distributions* (Russian) Teor. Verojatnost. i Primenen. **15** (1970), 469–497.

[Dob96] R. Dobrushin, *Perturbation methods of the theory of Gibbsian fields*, in R. Dobrushin, P. Groeneboom, M. Ledoux, *Lectures on probability theory and statistics*, Lectures from the 24th Saint-Flour Summer School held July 7–23, 1994, edited by P. Bernard, Lecture Notes in Mathematics **1648**, Springer, Berlin (1996), 1–66.

[DS85] R. L. Dobrushin, S. B. Shlosman, *Constructive criterion for the uniqueness of Gibbs field*, in J. Fritz, A. Jaffe and D. Szász (eds), *Statistical physics and dynamical systems*, papers from the second colloquium and workshop on random fields: rigorous results in statistical mechanics, held in Köszeg, August 26–September 1, 1984, Progress in Physics **10**, Birkhäuser, Boston (1985), 347–370.

[DS96] P. Diaconis, L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. **6** (1996), n° 3, 695–750.

[DZ98] A. Dembo, O. Zeitouni, *Large deviations techniques and applications*, second edition, Applications of Mathematics **38**, Springer, New York (1998).

[GD03] A. Granas, J. Dugundji, *Fixed point theory*, Springer Monographs in Mathematics, Springer, New York (2003).

[GH90] É. Ghys, P. de la Harpe, *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math. **83**, Birkhäuser (1990).

[GM83] M. Gromov, V. Milman, *A topological application of the isoperimetric inequality*, Amer. J. Math. **105** (1983), 843–854.

[Gri67] R. B. Griffiths, *Correlations in Ising ferromagnets III*, Commun. Math. Phys. **6** (1967), 121–127.

[Gro86] M. Gromov, *Isoperimetric inequalities in Riemannian manifolds*, in V. Milman, G. Schechtman, *Asymptotic theory of finite dimensional normed spaces*, Lecture Notes in Mathematics **1200**, Springer, Berlin (1986), 114–129.

[Jou07] A. Joulin, *Poisson-type deviation inequalities for curved continuous time Markov chains*, Bernoulli **13** (2007), n° 3, 782–798.

[Led01] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs **89**, AMS (2001).

[Lott] J. Lott, *Optimal transport and Ricci curvature for metric-measure spaces*, expository manuscript.

[LV] J. Lott, C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, preprint.

[Mar96a] K. Marton, *Bounding $\overline{\mathrm{d}}$-distance by informational divergence: a method to prove measure concentration*, Ann. Probab. **24** (1996), n° 2, 857–866.

[Mar96b] K. Marton, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. **6** (1996), n° 3, 556–571. Erratum in Geom. Funct. Anal. **7** (1997), n° 3, 609–613.

[Mar04] F. Martinelli, *Relaxation times of Markov chains in statistical mechanics and combinatorial structures*, in H. Kesten (ed), *Probability on discrete structures*, Encyclopaedia of Mathematical Sciences **110**, Springer, Berlin (2004), 175–262.

[MS86] V. Milman, G. Schechtman, *Asymptotic theory of finite dimensional normed spaces*, Lecture Notes in Mathematics **1200**, Springer, Berlin (1986).

[OEM] M. Hazewinkel, *The Online Encyclopaedia of Mathematics*, article *Wasserstein metric*, `http://eom.springer.de/W/w120020.htm`

[Oht07] S.-i. Ohta, *On the measure contraction property of metric measure spaces*, Comment. Math. Helv. **82** (2007), n° 4, 805–828.

[Oli] R. Imbuzeiro Oliveira, *On the convergence to equilibrium of Kac's random walk on matrices*, preprint, `arXiv:0705.2253`

[Oll04] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, Geom. Funct. Anal. **14** (2004), n° 3, 595–679.

[Oll07] Y. Ollivier, *Ricci curvature of metric spaces*, C. R. Math. Acad. Sci. Paris **345** (2007), n° 11, 643–646.

[OV00] F. Otto, C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), 361–400.

[Per] Y. Peres, *Mixing for Markov chains and spin systems*, lecture notes (2005).

[RS05] M.-K. von Renesse, K.-T. Sturm, *Transport inequalities, gradient estimates, and Ricci curvature*, Comm. Pure Appl. Math. **68** (2005), 923–940.

[Sam] M. D. Sammer, *Aspects of mass transportation in discrete concentration inequalities*, PhD thesis, Georgia institute of technology, 2005, `etd.gatech.edu/theses/available/etd-04112005-163457/unrestricted/sammer_marcus_d_200505_phd.pdf`

[Sch01] G. Schechtman, *Concentration, results and applications*, in *Handbook of the geometry of Banach spaces*, vol. 2, North-Holland, Amsterdam (2003), 1603–1634.

[Stu06] K.-T. Sturm, *On the geometry of metric measure spaces*, Acta Math. **196** (2006), n° 1, 65–177.

[Vil03] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics **58**, American Mathematical Society, Providence (2003).

[Vil08] C. Villani, *Optimal transport, old and new*, Grundlehren der mathematischen Wissenschaften **338**, Springer (2008).

Yann Ollivier, École normale supérieure de Lyon, 2009

Finding related pages using Green measures: An illustration with Wikipedia

Cette étude, avec Pierre Senellart, a été menée à l'automne 2006. Les résultats ont été présentés à l'édition 2007 de la conférence de l'Association for the Advancement of Artificial Intelligence (AAAI 2007), et publiés dans les actes de cette conférence, pp. 1427–1433.

Finding related pages using Green measures: An illustration with Wikipedia

Yann Ollivier & Pierre Senellart

Abstract

We introduce a new method for finding nodes semantically related to a given node in a hyperlinked graph: the Green method, based on a classical Markov chain tool. It is generic, adjustment-free and easy to implement. We test it in the case of the hyperlink structure of the English version of Wikipedia, the on-line encyclopedia. We present an extensive comparative study of the performance of our method versus several other classical methods in the case of Wikipedia. The Green method is found to have both the best average results and the best robustness.

1 Introduction

The use of tools relying on graph structures for extracting semantic information in a hyperlinked environment [Kle99] has had vast success, which led to a revolution in the search technology used on the World Wide Web [BP98]. In the same spirit, we present here a novel application of a classical tool from Markov chain theory, the Green measures, to the extraction of semantically related nodes in a directed graph. Such a technique can help a user find additional content closely related to a node i and thus guide her in the exploration of a graph. Google and Google Scholar both allow users to search for similar nodes, respectively in the Web graph and in the graph of scientific publications. This could also be useful in the case of the graph of an on-line encyclopedia like Wikipedia, where articles are seen as nodes of the graph and hyperlinks as edges between nodes: users are often interested in looking for articles on related topics, for instance to deepen their understanding of some concept. Other interests of an automatic method for finding related articles can be for instance to add missing links between articles [AdR05].

Our proposed method can be intuitively described as a PageRank [BP98] computation that continuously pours mass at node i. It is related to, but distinct from, so-called topic-sensitive PageRank [Hav03] (see below). The method provides a measure for similarity of nodes and could serve as a definition for some kind of "conceptual neighborhood" around i.

In order to be able to have a somewhat objective measure of the performance of the Green method, we compared it to several more classical approaches for extracting related pages in a graph. All these methods have been implemented, and tested on the graph of the English version of Wikipedia; though, to preserve generality of the approach, we did not implement any Wikipedia-specific tricks to enhance performance. A user study has been performed which allows us to evaluate and compare each of these techniques.

Our contributions are thus: (i) a novel use of Green measures to extract semantic information in a graph (ii) an extensive comparative study of the performance of different methods for finding related articles in the Wikipedia context. Note that we implemented "pure" versions of the methods: it is certainly easy to devise Wikipedia-specific enhancements to the methods, but we refused to do so in order to keep the comparison general. Even so, performance of the Green method was very satisfying.

We first introduce Green measures. Then we present different methods for extracting related nodes in a graph, based on Green measures, on PageRank, and on classical Information Retrieval approaches. The results of the experiment carried out to compare these methods are described next. Finally, we discuss related work and perspectives.

Additional data about the content presented here (including source code and full evaluation results) is available on the companion website for this paper [OS07].

2 Green measures

Notation for Markov chains. We collect here some standard facts and notation about Markov chains [Nor97].

Let (p_{ij}) be the transition probabilities of a Markov chain on a finite set of states X. That is, each p_{ij} is a non-negative number representing the probability to jump from node $i \in X$ to node $j \in X$; in particular, for each i we have $\sum_j p_{ij} = 1$. That is, the p_{ij}'s form a stochastic matrix.

For example, if X is given as a directed graph, we can define the *simple random walk on* X by setting $p_{ij} = 0$ if there is no edge from i to j, and $p_{ij} = 1/d_i$ if there is an edge from i to j, where d_i is the number of edges originating from i (if multiple edges are allowed, this definition can be adapted accordingly). This remark is very important since it allows one to view any hyperlinked environment as a Markov chain and to use and/or adapt Markov chain techniques.

A row vector $\mu = (\mu_i) : X \to \mathbb{R}$ indexed by X will be called a *measure on* X (negative values are allowed). The *(total) mass* of μ is $\sum \mu_i$. If moreover $\mu_i \geqslant 0$ and $\sum \mu_i = 1$, the measure will be called a *probability measure.*

We define the *forward propagation operator* M as follows: for any measure $\mu = (\mu_i)$ on X, the measure μM is defined by $(\mu M)_j := \sum_i \mu_i p_{ij}$, that is, each node i sends a part p_{ij} of its mass to node j. This corresponds to multiplication on the right by the matrix $M = (p_{ij})$, hence the notation μM. Note that forward propagation preserves the total mass $\sum \mu_i$.

Henceforth, we suppose, in a standard manner, that the Markov chain is irreducible and aperiodic [Nor97]. For the simple random walk on a graph, it amounts to the graph being strongly connected and the greatest common divisor of the lengths of all cycles being equal to 1.

Under all these assumptions, it is well-known that the Markov chain has a unique invariant probability measure ν, the *equilibrium measure*: that is, a unique measure ν with $\nu M = \nu$ and $\sum \nu_i = 1$. Moreover, for any measure μ such that $\sum \mu_i = 1$, the iterates μM^n converge to ν as $n \to \infty$. More precisely, the matrices M^n converge exponentially fast (in the number of iterations) to a matrix M^∞, which is of rank 1 and satisfies $M^\infty_{ij} = \nu_j$ for all i. The equilibrium measure ν can be thought of as a PageRank without random jumps on X [BP98].

Definition of Green measures. Green functions were historically introduced in electrostatic theory as a means of computing the potential created by a charge distribution; they have later been used in a variety of problems from analysis or physics [Duf01], and extended to discrete settings. The Green measure centered at i, as defined below, can really be thought of as the electric potential created on X by a unit charge placed at i [KSK66].

The *Green matrix* of a finite Markov chain is defined by

$$G := \sum_{t=0}^{\infty} (M^t - M^\infty)$$

where M^t is the t-th power of the matrix $M = (p_{ij})$, corresponding to t steps of the random walk. Since the M^t converge exponentially fast to M^∞, the series converges.

Now, for $i \in X$, let us define G_i, the *Green measure centered at* i, as the i-th row of the Green matrix G.

Let δ_i be the Dirac measure centered at i, that is, $\delta_{ij} := 1$ if $j = i$ and 0 otherwise. We have by definition $G_i = \delta_i G$. More explicitly, using that $M^\infty_{ij} = \nu_j$, we get

$$G_{ij} = \sum_{t=0}^{\infty} \left((\delta_i M^t)_j - \nu_j \right)$$

where $(\delta_i M^t)_j$ is of course the probability that the random walk starting at i is at j at time t. Since $\delta_i M^t$ is a probability measure and ν is as well, we see that for each i, G_i is a measure of total mass 0.

We now present other natural interpretations of the Green measures (in addition to electric potential).

PageRank with source at i. The sum

$$G_i = \sum_{t=0}^{\infty} (\delta_i - \nu) M^t$$

is a fixed point of the operator

$$\mu \mapsto \mu M + (\delta_i - \nu)$$

This fixed point is thus the equilibrium measure of a random walk with a source term δ_i which constantly pours a mass 1 at i, and a uniform sink term $-\nu$ (to preserve total mass). This is what makes Green measures different from PageRank and focused around a node.

This shows how Green measures can be computed in practice: Start with the row vector $\mu = \delta_i - \nu$ and iterate $\mu \mapsto \mu M + (\delta_i - \nu)$ some number of times. This allows to compute the Green measure centered at i without computing the whole Green matrix.

Time spent at a node. Since the equilibrium measure is ν, the average time spent at any node $j \in X$ by the random walk between steps 0 and t behaves, in the long run, like $(t+1)\nu_j$, whatever the starting node was. Knowing the starting node i leads to a correction term, which is precisely the Green measure centered at i. More precisely:

$$G_{ij} = \lim_{t\to\infty} \left(T_{ij}(t) - (t+1)\nu_j\right)$$

where $T_{ij}(t)$ is the average number of times the random walk starting at i hits node j between steps 0 and t (included).

Relationship with topic-sensitive PageRank. Topic-sensitive PageRank is a method for answering keyword queries on the World Wide Web which biases the PageRank method by focusing the PageRank computation around some *seed* subset of pages [Hav03]. It proceeds as follows. First, a list of topics is fixed, for each of which a list of *seed* Web pages is determined by hand. Second, for each different topic, a modified Markov chain is used which consists in, at each step, either following an outlink with probability $1 - c$, or jumping back to a seed page with probability c. Third, when answering queries, these modified PageRank values are combined with weights depending on the frequency of query terms in the seed documents.

Green measures are somewhat related to the modified Markov chain used as the second ingredient of topic-sensitive PageRank. Namely, let i be a single node that we use as the seed. Then the matrix $\tau(c)$ whose entry $\tau_{ij}(c)$ is the value at node j of the topic-sensitive PageRank with seed i is easily seen to be

$$\tau(c) = \sum_{t=0}^{\infty} c(1-c)^t M^t$$

where M is the transition matrix of the original Markov chain, and as above c is the rate at which the random walk jumps back to the seed. When c tends to 0, of course topic-sensitivity is lost, and the series tends to the matrix M^∞ all the rows of which are equal to the ordinary PageRank vector ν.

Now, we have:

$$\begin{aligned}\tau(c) - M^\infty &= \left(\sum_{t=0}^{\infty} c(1-c)^t M^t\right) - M^\infty \\ &= \sum_{t=0}^{\infty} c(1-c)^t \left(M^t - M^\infty\right)\end{aligned}$$

thanks to the identity $\sum c(1-c)^t = 1$. The Green matrix is thus related to the topic-sensitive matrix τ as follows:

$$G_{ij} = \lim_{c\to 0} \frac{1}{c} \left(\tau_{ij}(c) - \nu_j\right)$$

Thus, yet another interpretation of Green measures is as a way to get rid of the tendency of topic-sensitive PageRank to reproduce the global PageRank, and to extract meaningful information from it in a canonical way, without an arbitrary choice of the coefficient c.

3 Description of the methods

We now proceed to the definition of the five methods included in the evaluation: two Green-based methods and three more classical approaches.

The goal of each method is, given a node i in a graph (or in a Markov chain), to output an ordered list of nodes which are "most related" to i in some sense. All methods used here rely on scoring: given i, every node j is attributed a score $S^i(j)$. We then output the n nodes with highest score. Here we arbitrarily set $n = 20$, as we could not devise a natural and universal way to define a threshold.

3.1 Two Green-based methods

Green. The **Green** method relies directly on the Green measures described above. When looking for nodes similar to node i, compute the Green measure G_i centered at i. Now for each j, the value G_{ij} indicates how much node j is related to node i and can be taken as the score $S^i(j)$.

This score leads to satisfying results. However, nodes j with higher values of the equilibrium measure ν_j were slightly overrepresented. We found that performance was somewhat improved if an additional term favoring uncommon nodes j is introduced. Namely, we set

$$\boxed{S^i(j) := G_{ij} \log(1/\nu_j)}$$

The logarithmic term comes from information theory: $\log(1/\nu_j)$ is the quantity of information brought by the event "the random walk currently lies at node j", knowing that its prior probability is ν_j. This is very similar to the logarithmic term in the tf-idf formula used for **Cosine** below.

SymGreen. Since it mainly consists in following the Markov chain flow starting at node i, **Green** might miss nodes that point to i but are not pointed to by i, nodes which could be worth considering. The workaround is to symmetrize the Markov chain as follows: Given any Markov chain (p_{ij}) with stationary measure $\nu = (\nu_i)$, the *symmetrized Markov chain* is defined by

$$\tilde{p}_{ij} := \frac{1}{2}\left(p_{ij} + p_{ji}\frac{\nu_j}{\nu_i}\right)$$

which is still a stochastic matrix. This definition is designed so that the new Markov chain still has the same equilibrium measure ν. (Observe that simply forgetting edge orientation is not a proper way to symmetrize ν, since it will result in an invariant measure proportional to the degree of the node and ignore the actual values of the probabilities.)

This amounts to, at each step, tossing a coin and following the origin Markov chain either forward or backward (where the backward probabilities are given by $p_{ji}\nu_j/\nu_i$).

The Green measures $\tilde{G}_i$ for this new Markov chain $(\tilde{p}_{ij})$ can be defined in the same way, and as above the scores are given by

$$\boxed{S^i(j) := \tilde{G}_{ij}\,\log(1/\nu_j)}$$

3.2 PageRank-based methods

Arguably the most naive method for finding nodes related to a given node i is to look at nodes with high PageRank index in a neighborhood of i. Similar techniques are extensively used for finding *related pages* on the World Wide Web [Kle99, DH99]. Here by PageRank we mean the equilibrium measure of the random walk, that is, we discard random jumps (we set Google's *damping factor* to 1). Indeed, random jumps tend to spread the equilibrium measure more uniformly on a graph, whereas the goal here is to focus around a given node.

We describe two ways of using the equilibrium measure to identify nodes related to a given node.

PageRankOfLinks. The first method that springs to mind for identifying nodes related to i is to take the nodes pointed to by i and output them according to their PageRank.

Namely, let ν be the equilibrium measure of the random walk on the graph (or of the Markov chain). Let i be a node. The score of node j in the **PageRankOfLinks** method is defined by

$$\boxed{S^i(j) := \begin{cases} \nu_j & \text{if } \; p_{ij} > 0 \\ 0 & \text{if } \; p_{ij} = 0 \end{cases}}$$

LocalPageRank. Another PageRank-based method was implemented. It consists in, first, building a restricted graph centered at node i (namely, nodes obtained by following the links forwards, backwards, forwards-backwards and backwards-forwards), and then computing the equilibrium measure on this subgraph. The method outputs nodes of this subgraph, ordered according to this "local PageRank".

This method has an important flaw: As soon as the graph is highly connected, as is the case with Wikipedia, the neighborhood comprises a significant portion of the original graph. In such a case, the local equilibrium measure is very close to the global equilibrium measure, and so the results are not at all specific to i.

Due to its extremely poor results, this method was not included in the test. For example, on *Pierre de Fermat* the first 10 results in the output are *France*, *United States*, *United Kingdom*, *Germany*, *2005*, *2006*, *World War II*, *Italy*, *Europe* and *England*, showing no specific relationship to the base article but close to the global PageRank values.

3.3 Information retrieval-inspired methods

Standard information retrieval methods can be applied when only a graph/Markov chain is available, provided one is able to define the "content" of a node i. It is natural to interpret the set of nodes pointed to by i as the content of i, and moreover the transition probabilities p_{ij} can be thought of as the frequency of occurrence of j in i.

We tested two such methods: a cosine method using a tf-idf weight, and a cocitation index method.

Cosine with tf-idf weight. Cosine computations first use some transformation to represent each node/document in the collection by a vector in $\mathbb{R}^n$ for some fixed n. The proximity of two such vectors can then be measured by their cosine as ordinary vectors in $\mathbb{R}^n$ (or their angle, which amounts to the same as far as ordering is concerned).

One very usual such vector representation for documents is given by the *term frequency/inverse document frequency (tf-idf) weight* [SM84]. In our setting, it is adapted as follows.

Given a Markov chain defined by (p_{ij}) on a set of N elements (e.g. the random walk on a graph), for each node i the tf-idf vector x^i associated with i is an N-dimensional vector defined by

$$(x^i)_j := p_{ij} \log(N/d_j)$$

where d_j is the number of nodes pointing to j.

Cosine is then very simple: when looking for nodes related to node i, the score of node j is defined by

$$\boxed{S^i(j) := \cos(x^i, x^j)}$$

where x^i and x^j are seen as vectors in $\mathbb{R}^N$. Here, as usual, $\cos(x,y) = \frac{\sum x_k y_k}{\sqrt{\sum x_k^2}\sqrt{\sum y_k^2}}$.

Cocitations. A standard and straightforward method to evaluate document similarity is the cocitation index: two documents are similar if there are many documents pointing to both of them. This method, which originated in bibliometrics, is well-known and widely used for similar problems, see for instance [DH99] for an application to the Web graph.

In our context this simply reads as follows. When looking for nodes similar to a node i, the score of node j is given by

$$\boxed{S^i(j) := \#\{k,\, p_{ki} > 0 \text{ and } p_{kj} > 0\}}$$

Sometimes this method tends to favor nodes that have the same "type" as i rather than nodes semantically related to i but with a different nature. For example, when asked for pages related to *1989* (the year) in Wikipedia, the output is *1990*, *1991*... For the base article *Pierre de Fermat*, interestingly, it outputs several other great mathematicians.

4 Experimental results

In this section, we describe the experiments carried out to evaluate the performance of the methods presented above, on the graph of the English version of Wikipedia.

Graph extraction, implementation. A September 25th, 2006 dump of the English Wikipedia was downloaded from the URL `http://download.wikimedia.org/`. It was parsed in order to build the corresponding directed graph: nodes are the Wikipedia pages, and edges represent the links between pages. Multiple links were kept as multiple edges. Redirections (alternate titles for the same entry) were resolved. The most common templates (*Main*, *See also*, *Further*, *Details*...) were expanded. Categories (special entries on Wikipedia which are used to group articles according to semantic proximity, such as *Living people*) were kept as standalone pages, just as they appear on Wikipedia.

The resulting graph has $1,606,896$ nodes and $38,896,462$ edges; there are $73,860$ strongly connected components, the largest one of which contains $1,531,989$ nodes. We restrict ourselves to this strongly connected subgraph, in order to ensure convergence of computation of the equilibrium measure and Green measures.

Implementation of the methods is mostly straightforward, but here are a few caveats: 1. Because of the large size of the graph, memory handling must be considered with care; a large sparse graph library, relying on memory-mapped files, has been developed for this purpose. 2. Most methods require prior knowledge of the equilibrium measure for the graph, which is therefore computed once with very high accuracy. 3. Rather than the Green matrix, we compute the Green measure centered at i using the characterization of Green measures as fixed point of the operator $\mu \mapsto \mu M + (\delta_i - \nu)$.

The computation time for **Green** is less than 10s per article on a 3GHz desktop PC; that of **SymGreen** is typically between 15s and 30s. The other methods range from a few seconds to three minutes (**Cosine**). Computation of **Green** is easily parallelizable; we estimate that computation of the full Green matrix would take less than two weeks on a 10 PC cluster, after which the answers are instantaneous.

Evaluation methodology. We carried out a blind evaluation of the methods on 7 different articles, chosen for their diversity: (i) *Clique (graph theory)*: a very short, technical article. (ii) *Germany*: a very large article. (iii) *Hungarian language*: a medium-sized, quite technical article. (iv) *Pierre de Fermat*: a short biographical article. (v) *Star Wars*: a large article, with an important number of links. (vi) *Theory of relativity*: a short introductory article pointing to more specialized articles. (vii) *1989*: a very large article, containing all the important events of year 1989. It

Table 1: Output of **Green** on the articles used for evaluation.

Clique (graph theory)	*Germany*	*Hungarian language*	*Pierre de Fermat*	*Star Wars*	*Theory of relativity*	*1989*
1. Clique (graph theory) 2. Graph (mathematics) 3. Graph theory 4. Category: Graph theory 5. NP-complete 6. Complement graph 7. Clique problem 8. Complete graph 9. Independent set 10. Maximum common subgraph isomorphism problem 11. Planar graph 12. Glossary of graph theory 13. Mathematics 14. Connectivity (graph theory) 15. Computer science 16. David S. Johnson 17. Independent set problem 18. Computational complexity theory 19. Set 20. Michael Garey	1. Germany 2. Berlin 3. German language 4. Christian Democratic Union (Germany) 5. Austria 6. Hamburg 7. German reunification 8. Social Democratic Party of Germany 9. German Empire 10. German Democratic Republic 11. Bavaria 12. Stuttgart 13. States of Germany 14. Munich 15. European Union 16. National Socialist German Workers Party 17. World War II 18. Jean Edward Smith 19. Soviet Union 20. Rhine	1. Hungarian language 2. Slovakia 3. Romania 4. Slovenia 5. Hungarian alphabet 6. Hungary 7. Croatia 8. Category: Hungarian language 9. Turkic languages 10. Finno-Ugric languages 11. Austria 12. Serbia 13. Uralic languages 14. Ukraine 15. Hungarian grammar (verbs) 16. German language 17. Hungarian grammar 18. Khanty language 19. Hungarian phonology 20. Finnish language	1. Pierre de Fermat 2. Toulouse 3. Fermat's Last Theorem 4. Diophantine equation 5. Fermat's little theorem 6. Fermat number 7. Grandes écoles 8. Blaise Pascal 9. France 10. Pseudoprime 11. Lagrange's four-square theorem 12. Number theory 13. Fermat polygonal number theorem 14. Holographic will 15. Diophantus 16. Euler's theorem 17. Pell's equation 18. Fermat's theorem on sums of two squares 19. Fermat's spiral 20. Fermat's factorization method	1. Star Wars 2. Dates in Star Wars 3. Palpatine 4. Jedi 5. Expanded Universe (Star Wars) 6. Star Wars Episode I: The Phantom Menace 7. Star Wars Episode IV: A New Hope 8. Obi-Wan Kenobi 9. Star Wars Episode III: Revenge of the Sith 10. Coruscant 11. Anakin Skywalker 12. Lando Calrissian 13. Luke Skywalker 14. Star Wars: Clone Wars 15. List of Star Wars books 16. George Lucas 17. Star Wars Episode II: Attack of the Clones 18. Splinter of the Mind's Eye 19. List of Star Wars comic books 20. The Force (Star Wars)	1. Theory of relativity 2. Special relativity 3. General relativity 4. Spacetime 5. Lorentz covariance 6. Albert Einstein 7. Principle of relativity 8. Electromagnetism 9. Lorentz transformation 10. Inertial frame of reference 11. Speed of light 12. Galilean transformation 13. Local symmetry 14. Category: Relativity 15. Galilean invariance 16. Gravitation 17. Global symmetry 18. Tensor 19. Maxwell's equations 20. Introduction to general relativity	1. 1989 2. Cold War 3. 1912 4. Tiananmen Square protests of 1989 5. Soviet Union 6. German Democratic Republic 7. George H. W. Bush 8. 1903 9. Communism 10. 1908 11. 1929 12. Ruhollah Khomeini 13. March 1 14. Czechoslovakia 15. June 4 16. The Satanic Verses (novel) 17. 1902 18. November 7 19. October 9 20. March 14
Mark: 7.6/10	Mark: 7.0/10	Mark: 6.2/10	Mark: 7.3/10	Mark: 7.4/10	Mark: 8.1/10	Mark: 5.4/10

Table 2: Output of **SymGreen**, **Cosine**, **Cocitations**, and **PageRankOfLinks** on sample articles.

SymGreen		**Cosine**		**Cocitations**		**PageRankOfLinks**	
Pierre de Fermat	*Germany*	*Pierre de Fermat*	*Germany*	*Pierre de Fermat*	*Germany*	*Pierre de Fermat*	*Germany*
1. Pierre de Fermat 2. Mathematics 3. Probability theory 4. Fermat's Last Theorem 5. Number theory 6. Toulouse 7. Diophantine equation 8. Blaise Pascal 9. Fermat's little theorem 10. Calculus	1. Germany 2. Berlin 3. France 4. Austria 5. German language 6. Bavaria 7. World War II 8. German Democratic Republic 9. European Union 10. Hamburg	1. Pierre de Fermat 2. ENSICA 3. Fermat's theorem 4. International School of Toulouse 5. École Nationale Supérieure d'Électronique, d'Électrotechnique... 6. Languedoc 7. Hélène Pince 8. Community of Agglomeration of Greater Toulouse 9. Lilhac 10. Institut d'études politiques de Toulouse	1. Germany 2. History of Germany since 1945 3. History of Germany 4. Timeline of German history 5. States of Germany 6. Politics of Germany 7. List of Germany-related topics 8. Hildesheimer Rabbinical Seminary 9. Pleasure Victim 10. German Unity Day	1. Pierre de Fermat 2. Leonhard Euler 3. Mathematics 4. René Descartes 5. Mathematician 6. Gottfried Leibniz 7. Calculus 8. Isaac Newton 9. Blaise Pascal 10. Carl Friedrich Gauss	1. Germany 2. United States 3. France 4. United Kingdom 5. World War II 6. Italy 7. Netherlands 8. Japan 9. 2005 10. Category: Living people	1. France 2. 17th century 3. March 4 4. January 12 5. August 17 6. Calculus 7. Lawyer 8. 1660 9. Number theory 10. René Descartes	1. United States 2. United Kingdom 3. France 4. 2005 5. Germany 6. World War II 7. Canada 8. English language 9. Japan 10. Italy
Mark: 7.0/10	Mark: 5.5/10	Mark: 2.9/10	Mark: 7.4/10	Mark: 5.4/10	Mark: 2.1/10	Mark: 2.5/10	Mark: 1.1/10

was unreasonable to expect our testers to evaluate more articles. In order to avoid any bias, we did not run the methods on these 7 articles before the evaluation procedure was launched.

People were asked to assign a mark between 0 and 10 (10 being the best) to the list of the first 20 results returned by each method on these articles, according to their relevance as "related articles" lists. Each evaluator was free to interpret the meaning of the phrase "related articles". The lists were unlabeled, randomly shuffled, and in a potentially different order for each article. The evaluators were allowed to skip articles they did not feel confident enough to vote on. There has been a total of 67 participants, which allows for reasonably good confidence intervals.

Performance of the methods. Table 1 shows the output of **Green** on each evaluated article. Due to lack of space, we only present a portion of the outputs of the other methods in Table 2. The full output and detailed evaluation results can be found in [OS07].

The average marks given by the evaluators are presented in a radar chart on Figure 1. Each axis stands for the mark given for an article: from worst (0/10) at the

Table 3: Evaluation results. For each method, the following figures are given: average mark, averaged on all articles; 90% Student's t-distribution confidence interval; article-to-article standard deviation; evaluator-to-evaluator standard deviation; global count of 10/10 marks; average mark for each article.

	Green	SymGreen	Cosine	Cocitations	PageRankOfLinks
Average mark	7.0	6.3	5.2	4.5	2.2
90% confidence interval	±0.3	±0.3	±0.3	±0.3	±0.2
Article std. dev.	0.9	1.3	2.2	1.9	2.0
Evaluator std. dev.	1.7	1.7	1.9	2.0	1.6
Number of 10/10	25	10	12	9	4
Clique (graph theory)	7.6	7.6	6.2	7.5	6.8
Germany	7.0	5.5	7.4	2.1	1.1
Hungarian language	6.2	5.8	3.3	3.8	0.5
Pierre de Fermat	7.3	7.0	2.9	5.4	2.5
Star Wars	7.4	6.9	7.8	4.7	0.6
Theory of relativity	8.1	7.7	6.7	6.1	2.7
1989	5.4	3.8	2.1	1.9	1.1

Figure 1: Radar chart of the average marks given to each method on the various base articles.

center to best (10/10) at the periphery, while each row represents a method (cf. the legend). Table 3 gives global statistics about the performance of the methods.

Absolute marks should be taken with caution: it is probable that a human-designed list of related pages would not score very close to 10/10, but maybe closer to 8/10. Indeed, the evaluator-to-evaluator standard deviation for a given article is always between 1.5 and 2.0. For example, on *Theory of relativity*, **Green** gets 8.1/10 though it was attributed a top 10/10 mark by a significant number of evaluators, including several experts in this field.

Green presents the best overall performance. The difference between global scores of **Green** and of the best classical approach, **Cosine**, is 1.8, which is statistically significant. **Green** comes out first for all but two articles, where it is second with a hardly significant gap (0.4 in both cases). Moreover, **Green** is extremely robust: First, it has a low article-to-article standard deviation, and a look at Figure 1 shows that it never performs very badly. Second, there are very few irrelevant words in its output, as can be seen on Table 1; the high number of 10/10 given to **Green** is perhaps a measure of this fact. Finally, some of the related articles proposed by **Green** are both highly semantically relevant and completely absent from the output of other methods: this is the case of *Finnish language* for *Hungarian language* (linguists now consider both languages closely related), and of *Tiananmen Square protests* or *The Satanic Verses* for *1989*.

SymGreen presents a profile similar to **Green** for both performance and robustness. Actually, though its overall mark is slightly less on the evaluated articles, on other articles we experimented with in an informal way, it seems more robust than **Green**. It might in fact be better adapted for other contexts, especially in less highly connected graphs.

Cosine performs best of the "classical" methods, but is clearly not as good as the Green-based ones. Both very good and very bad performance occur: compare for instance *Germany* and *Pierre de Fermat* in Table 2. Thus, this method is unstable, which is visible in its high article-to-article standard deviation. Moreover, even in the case when it performs well, as for *Germany*, completely irrelevant or anecdotal entries are proposed, like *Pleasure Victim* or *Hildesheimer Rabbinical Seminary*. Testing the methods informally on more articles confirmed this serious instability of **Cosine**.

Cocitations does not give very good results, but it is still interesting: more than *related* articles, it outputs lists of articles of the same *type*, giving for instance names of great mathematicians of the same period for *Pierre de Fermat*, languages for *Hungarian language* or years for *1989*.

PageRankOfLinks is the worst of the methods tested (although **LocalPageRank**, not formally tested here, is even worse). It basically outputs variations on the global PageRank values whatever the base article, except on articles with very few links.

5 Related work

To our knowledge, this is the first use of discrete Green measures in the field of information retrieval on graphs or hyperlinked structures.

The relationship between Green measures and topic-sensitive PageRank [Hav03] has been discussed above. Note that, in addition to the mathematical differences, the purpose is not the same: in the case of topic-sensitive PageRank, classical keyword Web search focused on a specific part of the Web, with a measure of topic-wise importance; in our case, a measure of similarity unmarred by global PageRank values, and a definition of *conceptual neighborhoods* in a graph.

The problem of finding related nodes on the World Wide Web is not new. In his original well-known paper about *hubs* and *authorities* [Kle99], Kleinberg suggests using authorities in a focused subgraph in order to compute *similar-page queries*; apart from the use of authorities instead of PageRank, this is very similar to **LocalPageRank**, which performs poorly on Wikipedia. In [DH99], the authors present two different approaches for finding related pages on the Web: the *Companion* algorithm, which uses authorities scores in a local subgraph, and a cocitation-based algorithm.

In the specific case of Wikipedia, [AdR05] uses a cocitation approach to identify missing links. We saw that **Cocitations** fared much worse than **Green** in our experiment. *Synarcher* [Kri05] is a program for synonym extraction in Wikipedia, relying on authority scores in a local subgraph (comparable to **LocalPageRank**) together with the information provided by Wikipedia's category structures. In [GB05] a technique is presented to modify a classical text mining similarity measure (based on full textual content) by taking the hyperlinks into account using machine learning; no application to the problem of finding related pages is given.

6 Conclusion and perspectives

We showed how to use Green measures for the extraction of related nodes in a graph. This is a generic, parameter-free algorithm, which can be applied *as is* to any directed graph. We have described and implemented in a uniform way other classical approaches for finding related nodes. Finally, we have carried out a user study on the example of the graph of Wikipedia. The results show that the Green method has three advantages: 1. Its average performance is high, significantly above that of all other methods. 2. It is *robust*, never showing a bad performance on an article. 3. It is able to unveil semantic relationships not found by the other methods.

There is much room for extensions and improvements, either on the theoretical or the application side. For example it is easy to design variations on the Green method using standard variations on PageRank, such as HITS [Kle99]. Also, there is a continuous interpolation between **Green**, which follows only forward links, and **SymGreen**, which is bidirectional and tends to broaden the range of results (and is probably more robust). This could be used as a "specificity/generality" cursor.

A strong point of the methods presented here is that they rely only on the graph structure. It is very likely that, in the specific case of Wikipedia, we can improve performance by taking into account the textual content of the articles, the categories, some templates... although the raw method already performs quite well.

An obvious application is to try the Green method on the Web graph; this requires much more computational power, but seems feasible with large clusters of PCs. More generally, the method could be directly applied to any other context featuring associative networks.

7 Acknowledgments

We would like to thank all 67 participants who took the time to evaluate the different methods, and Serge Abiteboul for his feedback on the topic.

References

[AdR05] Sisay Fissaha Adafre and M. de Rijke. Discovering missing links in Wikipedia. In *Workshop on Link Discovery: Issues, Approaches and Applications*, Chicago, USA, August 2005.

[BP98] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[DH99] Jeffrey Dean and Monika Rauch Henzinger. Finding related pages in the World Wide Web. *Computer Networks*, 31(11–16):1467–1479, 1999.

[Duf01] Dean G. Duffy. *Green's functions with applications*. Studies in Advanced Mathematics. Chapman & Hall/CRC, Boca Raton, FL, 2001.

[GB05] David Grangier and Samy Bengio. Inferring document similarity from hyperlinks. In *Conference on Information and Knowledge Management*, Bremen, Germany, November 2005.

[Hav03] Taher H. Haveliwala. Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):784–796, 2003.

[Kle99] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[Kri05] Andrew A. Krizhanovsky. Synonym search in Wikipedia: Synarcher. `arXiv:cs.IR/0606097`, 2005.

[KSK66] John G. Kemeny, J. Laurie Snell, and Anthony W. Knapp. *Denumerable Markov chains*. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1966.

[Nor97] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1997.

[OS07] Yann Ollivier and Pierre Senellart. Finding related pages using Green measures: An illustration with Wikipedia, companion website. `http://pierre.senellart.com/wikipedia/`, 2007.

[SM84] Gerard Salton and Michael McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill Book Company, 1984.

Rate of convergence of crossover operators

Le texte suivant est mon premier article, écrit juste après mon mémoire de DEA, suite à ma participation à un groupe de travail sur les algorithmes génétiques organisé par Raphaël Cerf à Orsay en 1999–2000. Il est paru au volume 23, n°1 (2003) de Random Structures and Algorithms*, pp. 58–72.*

Rate of convergence of crossover operators

Yann Ollivier

Abstract

We study the convergence of mating operators on $\{0,1\}^n$. In particular, we answer questions of Rabani, Rabinovich and Sinclair (*cf.* [5]) by giving tight estimates on the divergence between the finite- and infinite-population processes, thus solving positively the problem of the simulability of such quadratic dynamical systems.

Introduction, main results

We study from a theoretical point of view the rate of convergence of a mating operator between two "genomes", in the framework of population genetics or genetic algorithms: a population is made up of individuals defined by a genome, which is a string of symbols (taken in $\{0,1\}$ for convenience).

The mating operator consists in having a stem population replaced by a new one in the following way: an individual from the new population is obtained by randomly, uniformly sampling two distinct individuals in the stem population and mixing their genomes in some prescribed way. These operations are repeated independently in order to obtain all the individuals of the new population.

Intuitively, mating seems to mix the genes present in the stem population. Biology handbooks claim that the interest of sexual reproduction is to keep a high level of diversity and to mix all available genes. Thus, it can be interesting to study the speed of such a mixing.

We choose a genome length n, and we define the random offspring of a mating between two elements of $\{0,1\}^n$ as follows: Fix a probability distribution Π (a *crossover operator*) on the set of subsets of $\{1\dots n\}$. Sample an $S \subset \{1\dots n\}$ from Π. Then, the offspring of the pair $x, y \in \{0,1\}^n$ is a random element $z \in \{0,1\}^n$ whose i-th bit z_i is equal to x_i if $i \in S$, or y_i if $i \notin S$.

According to the chosen distribution Π, different kinds of mating can be obtained. The simplest one is *uniform crossover*: Π is the uniform distribution on all subsets of $\{1\dots n\}$. This amounts to choosing each of the bits z_i to be equal to x_i or y_i independently of each other with probability 1/2.

We consider a finite population process of size k: at any step, the population is made up of k (not necessarily distinct) elements of $\{0,1\}^n$. The population for the next step is obtained by uniformly picking, k times with replacement, a random pair of distinct individuals[1] in the previous population, by having them generate a child from our mating operator and by putting the child in the new population.

Let π_t be the random k-tuple in $\{0,1\}^n$ obtained after t iterations of the process, given an initial k-tuple π_0.

We want to compare this process with the so-called "infinite-population process" where an infinite population is a probability distribution on $\{0,1\}^n$: the law of an element from the distribution p_{t+1} is obtained by sampling two individuals according to p_t and mating them according to Π. For a given p_0, we obtain a (deterministic) sequence p_t of probability distributions on $\{0,1\}^n$.

The infinite-population process is fairly well-known (see the work by Y. Rabani, Y. Rabinovich and A. Sinclair in [5]). It converges to a distribution p_∞ which depends on p_0 in the following way: under p_∞, the bits of an individual are chosen independently of each other, and their value is 0 or 1 with the same probability as in p_0. In other words, the proportion, in the population, of 0 and 1 at each position in the genome is invariant under the process, but the values at different positions tend to be independent.

The authors of [5] give essentially tight upper and lower bounds on the convergence of the infinite-population process. Let us recall their main result.

Definition 1.
Let the distance $\left|p-p'\right|$ between two probability distributions p and p' be

$$\left|p-p'\right| = \frac{1}{2}\sum_{x\in\{0,1\}^n}\left|p(x)-p'(x)\right| = \sup_{X\subset\{0,1\}^n}\left|p(X)-p'(X)\right| \leqslant 1$$

Then, under a natural non-degeneracy assumption on the mating operator, we have $|p_t-p_\infty| \leqslant n^2\, r_\Pi^t$, where $r_\Pi < 1$ is a constant depending on the crossover operator (equal to $1/2$ for uniform crossover). Furthermore, these authors show that for particular crossover operators, this result is essentially tight, in the sense that e.g. for uniform crossover, the time required for $|p_t-p_\infty|$ to be less than $1/4$ (the "mixing time") is at least $\log_2 n - O(1)$ for some initial population. At the end of the paper we prove a similar but different tightness result (see section 2.A).

On the other hand, the finite-population process is harder to comprehend. It can be thought of as an approximation of the infinite-population process; but it seems that, in order to determine an individual at some step, it would be necessary to know its two parents, its four grandparents, …its 2^t forefathers. Thus if the population is small, some forefathers will appear several times in the family tree, which will result in undesired correlations.

[1] By "distinct" individuals we do not mean that their genomes are necessarily distinct, but that they correspond to distinct indices $i,j \leqslant k$ in the population. This assumption is natural for the modelling of sexual reproduction in biological systems (excluding occasional parthenogenesis). If we release this assumption, all results stated here remain true with the constant r_Π replaced with $r_\Pi(1-1/k)+1/k$.

This problem arises for all so-called "quadratic dynamical systems" (*cf.* [6]), when we are given some random "mating" between two individuals in a given space, and we evolve probability measures on this space by defining the law of an individual at time $t+1$ to be the law of the offspring of two individuals picked from the law at time t. The difficulty of simulating a quadratic dynamical system has been formalized (*cf.* [1]): indeed, such systems can solve in polynomial time any PSpace problem.

The comparison between the two processes goes as follows: Given an infinite population p_0, we sample k individuals from it. This results in a random k-tuple π_0. This k-tuple evolves as described above, and we denote by π_t the k-tuple at time t.

Actually, π_t seen as a probability measure on $\{0,1\}^n$ (each element of the k-tuple having weight $1/k$) is of course not a good approximation of the infinite population p_t since it is supported on only k individuals, whereas in general p_t is supported on all of $\{0,1\}^n$ more or less uniformly.

We could rather try to compare the law of the random k-tuple π_t with the law $p_t^{\otimes k}$ of a random k-sample from p_t (after all, π_0 was a k-sample from p_0). As it turns out, this is not a good comparison. Indeed, after some time, π_t is very probably made up of k clones of one single individual (this is because at each step, with small probability, some genetic information gets lost). This well-known phenomenon is termed *coalescence*. (By the way, this shows that the process π_t converges.) We will return to this in section 3.

But the random individual making up this uniform population π_t will not always be the same, and its probability law will be close to p_t, which is what we wish. Thus, the law of a single element (e.g. the first one) of π_t, taken alone, is a good approximation to p_t.

Hence, denote by q_t the probability law of the first element of the random k-tuple π_t.

Y. Rabani, Y. Rabinovich and A. Sinclair prove that $|q_t - p_t| \leqslant \dfrac{4n^2 t}{k}$. Our main result is that

$$|q_t - p_\infty| \leqslant \frac{n^2}{C_\Pi\, k} + n^2\, r_\Pi^t$$

where r_Π is the same constant depending on the crossover operator as in Y. Rabani, Y. Rabinovich and A. Sinclair's result on the infinite-population process, and $C_\Pi = 1 - r_\Pi + 1/k$.

For example, for uniform crossover, this leads to

$$|q_t - p_\infty| \leqslant n^2 \left(\frac{2}{k+2} + \frac{1}{2^t} \right)$$

Considering uniform crossover, letting $k \to \infty$ so that the finite-population process closely follows the infinite-population process, and applying our lower bound stated above in that case, shows that the term $n^2/2^t$ (with $r_\Pi = 1/2$) is tight up to a $O(n)$ factor.

Furthermore, we prove that for k big enough, for some initial population, we have $|p_\infty - q_\infty| \geqslant \dfrac{n}{Ck}$ for some constant $C \leqslant 32$. So, our bounds are essentially tight up to replacement of n^2 by n, which affect the mixing time by at most a factor of 2.

At the end of the paper, we give a proof of similar results regarding mean-time (before coalescence) approximation of a whole population rather than a single individual.

1 Convergence of the finite population process

1.A Background : convergence of the infinite population process

We recall here the results of Y. Rabani, Y. Rabinovich and A. Sinclair (*cf.* [5]).

Let p_0 be a probability distribution on $\{0,1\}^n$. Let a_{i0} be the probability that the i-th bit of an individual sampled from p_0 is 0, and $a_{i1} = 1 - a_{i0}$.

Denote by p_∞ the probability law which, to the individual $x = x_1 x_2 \dots x_n$, assigns the weight $p_\infty(x_1 x_2 \dots x_n) = \prod a_{ix_i}$. This is the probability law where each bit equals 0 or 1 with the same probability as in p_0, but where different bits are chosen independently of each other.

For example, if p_0 is the distribution that puts weight $1/2$ on the individual $000\dots0$ and $1/2$ on $111\dots1$, then p_∞ is the uniform distribution on $\{0,1\}^n$.

Here, Y. Rabani, Y. Rabinovich and A. Sinclair make a non-degeneracy assumption on the chosen crossover operator Π: they demand that each two different positions $1 \leqslant i, j \leqslant n$ have a positive probability to be separated by the crossover, that is, that there be an $S \subset \{1\dots n\}$ with $\Pi(S) > 0$ and $i \in S, j \notin S$ (otherwise, these two positions could be considered as one single two-bit block).

This natural assumption holds for all usual crossovers. The authors are especially interested in the following cases:

- Uniform crossover: Π is the uniform distribution on subsets of $\{1\dots n\}$, each bit is picked independently from one of the two parents.
- One-point crossover: Choose a position $1 \leqslant i \leqslant n+1$ uniformly. Those bits with position less than i will be picked from one parent and the other bits from the other one. So Π gives equal weight to the $n+1$ sets $\varnothing$, $\{1\}$, $\{1,2\}$, $\{1,2,3\}$, $\dots$, $\{1,2,\dots,n\}$.
- Poisson crossover: We begin at position 0, picking successive bits of one parent. Then after some time we jump to the other parent and pick some successive bits from it, etc. At each step, the probability to jump from one parent to the other is the same.

Under the non-degeneracy assumption, [5] states the following result:

Theorem 2[5].
The infinite-population process p_t converges to p_∞ (as probability measures on $\{0,1\}^n$).

Furthermore, they give a good estimate of the rate of convergence. This depends on details of the crossover operator. Following their notation, let $r_{ij}(\Pi)$ be the probability that positions i and j are *not* separated by an $S \subset \{1 \dots n\}$ sampled from Π. Let $r_\Pi = \max_{i,j} r_{ij}(\Pi)$. The non-degeneracy assumption states that $r_\Pi < 1$.

Then

Theorem 3[5].
The distance between the population at time t and the limit population p_∞ satisfies

$$|p_t - p_\infty| \leqslant n^2\, r_\Pi^t$$

For instance, $r_\Pi = 1/2$ for uniform crossover, and hence $|p_t - p_\infty| \leqslant n^2/2^t$.

1.B Convergence for finite populations

Recall that q_t is the law of the first element of the random k-tuple π_t after t steps of the finite-population process, when π_0 is made up of k independent samplings from p_0.

In [5], Y. Rabani, Y. Rabinovich and A. Sinclair show that

$$|q_t - p_t| \leqslant \frac{4n^2 t}{k}$$

for any crossover operator. We show here, using similar techniques, that

Theorem 4.

$$|q_t - p_\infty| \leqslant \frac{n^2}{k\,(1 - r_\Pi + 1/k)} + n^2\, r_\Pi^t$$

with r_Π as above.

In particular, $|q_\infty - p_\infty| \leqslant n^2/(k(1 - r_\Pi + 1/k))$, and for $k \gg n^2$, the mixing time is less than $2 \log_{1/r_\Pi} n$.

For the sake of optimality, we show below (section 2.B) that $|q_\infty - p_\infty| \geqslant n/Ck$ in some cases (where C is a constant). This is an intrinsic bias due to the finite population approximation.

Note that this bound is *not* obtained from bounding $|q_t - p_t|$ and then using the bound on $|p_t - p_\infty|$. We directly compare q_t with p_∞. The bias of p_t and of q_t compared to p_∞ may not be of the same kind.

Proof.
We look at the process π_t in the following way: To generate π_t, we first leave π_0 unspecified, we choose a family tree from generation 0 to generation t from the correct probability distribution, and, fully independently, we fill π_0 by sampling k individuals from p_0. Then we look at how the bits of generation 0 propagate through the tree.

More precisely, a "family tree" is a structure in which, for each $t \geqslant 1$ and for each individual number i in generation t, two distinct members i_1 and i_2 of the previous generation are specified, together with a "mask" $S \subset \{1 \dots n\}$ describing those bits of i that come from i_1 or i_2. This tree gets a probability, which is the product of the probabilities, under Π, of all masks appearing in it, divided by $(k(k-1))^{kt}$ which corresponds to all possible choices of the parents of all individuals.

Once a family tree is given, we fill the bits of generation 0 using the distribution p_0, independently of this tree. Under these conditions, we are in a position to travel back through the tree and tell, for each bit of any individual at generation t, which bit from which individual of generation 0 it comes from.

We then note that, if we get a tree such that all n bits of the first individual of generation t come from *distinct* individuals from generation 0, these n bits come from n individuals independently sampled from p_0. The values, 0 or 1, of these bits are thus independent, and the i-th bit is a 1 with probability a_{i1} (in our earlier notation). In other words, if we get a tree where the n bits of the first individual of π_t come from distinct individuals, then the law of this individual is exactly p_∞ and we are done.

Then, a little manipulation of the definition of $|q_t - p_\infty|$ shows that this distance is less than the probability that the sampled tree be not of the above kind.

Let's evaluate this probability. Consider two bits of the first individual at generation t. If at some time $t' \leqslant t$, these bits belong to the same individual, the probability that they come from the same parent of this individual at time $t'-1$ is a number p depending on Π, with $p \leqslant r_\Pi$. If at time t' they belong to two different individuals, their respective parents are chosen independently in $\pi_{t'-1}$, and the probability that they come from the same individual of $\pi_{t'-1}$ is $1/k$.

Going back through the tree, we thus have a Markov chain with the following transition probabilities between the two states D (the two bits belong to two distinct individuals) and S (they belong to the same individual): $D \to D$ with probability $1-1/k$, $D \to S$ with probability $1/k$, $S \to S$ with probability $p \leqslant r_\Pi$, $S \to D$ with probability $1-p$.

A (very simple) calculation gives that, knowing that at time t the bits are together, the probability to get a family tree where these two bits are together at time 0 is

$$\frac{1}{k(1-p+1/k)} + \left(p - \frac{1}{k}\right)^t \left(1 - \frac{1}{k(1-p+1/k)}\right)$$

which, since $p \leqslant r_\Pi$, is less than

$$\frac{1}{k(1-r_\Pi+1/k)} + \max(r_\Pi, 1/k)^t$$

In general, $1/k$ will be smaller than r_Π. If not, note that $\dfrac{1}{k(1-r_\Pi+1/k)} + \dfrac{1}{k^t} \leqslant \dfrac{2}{k(1-r_\Pi+1/k)}$ as soon as $k \geqslant 2, t \geqslant 2$ (the cases $k=1$ or $t=0,1$ being trivial). Anyway, the probability in question is less than $\dfrac{2}{k(1-r_\Pi+1/k)} + r_\Pi^t$.

This was for one pair of bits of the first individual of generation t. There are $n(n-1)/2$ such pairs. The probability that the sampled tree presents two bits with the same ancestor from generation 0 is, then, less than $\dfrac{n(n-1)}{2}\left(\dfrac{2}{k(1-r_\Pi+1/k)}+r_\Pi^t\right)$, hence the theorem. □

The main difference with the analysis in [5] is that we make a more refined analysis of collisions: collisions are not so much disturbing, as two bits which collide at some time can be separated again further back in the tree. Note that this leads to a comparison of q_t to p_∞ and not to p_t, because once a collision has occurred the correlation between q_t and p_t is lost, and further separation of the collided bits does not restore this correlation which relies on the specific structure of the tree.

2 Lower bounds on convergence

We now turn to proving that the bounds for convergence obtained so far are essentially tight. Results in this direction for infinite populations already appear in [5]. We give below a tightness result for finite populations. As a template, we begin by giving a tightness result for uniform crossover in infinite populations which is different from that of [5].

2.A Lower bound for uniform crossover in infinite populations

Recall Theorem 3: $|p_t - p_\infty| \leqslant n^2\, r_\Pi^t$. The asymptotic part (in t) of this is tight: indeed, there exists a population p_0 such that for all t, $|p_t - p_\infty| \geqslant r_\Pi^t/2$.

Define the mixing time τ of the process as the smallest t such that whatever the initial population p_0 was, we have $|p_t - p_\infty| < 1/4$. So $\tau \leqslant 2\log_{1/r_\Pi} n + 2\log_{1/r_\Pi} 2$, which is a fairly good result.

The authors of [5] show that for particular crossover operators, this result is essentially tight. Their argument depends on the details of the crossover. For instance, for uniform crossover, they obtain $\tau \geqslant \log_2 n - O(1)$; hence the bound on the mixing time is tight up to a factor of 2. For Poisson crossover, their result is tight up to a factor of $O(\log\log\log n)$.

We prove that for uniform crossover, the result is essentially tight in a different sense than that of [5]. Namely, we show that for some initial population p_0, we have $|p_t - p_\infty| \geqslant \dfrac{n}{C\,2^t}$ for some constant C, for t large enough. So we cannot replace n^2 by an expression smaller than n in the upper bound above for $|p_t - p_\infty|$.

These results are not directly comparable: the one deals with the time required to reach some threshold, whereas the other reflects the asymptotic behavior. However, assuming that our estimate of the asymptotic behavior is tight even for short times would result in the same estimate $\log_2 n - O(1)$ for the mixing time.

Theorem 5.
For uniform crossover, for n and t large enough, for some initial population p_0, we have

$$|p_t - p_\infty| \geqslant \frac{n}{32\cdot 2^t}$$

Inspecting the proof reveals that the result holds as soon as $n \geqslant 8$ and $t \geqslant 3\log_2 n + 4$ (the time from which the theorem holds depends inevitably on n, since otherwise $n/(32\ 2^t)$ could be greater than 1).

Proof.
Let us have a fresh look at how an individual from generation t is built. First, let's fix the 2^t ancestors of this individual at time 0, sampled from p_0. Then, we observe that, under uniform crossover, each of the n bits of the individual comes from one of these ancestors, which we will call the ancestor of the specified bit. In the case of uniform crossover, by a straightforward induction, the ancestor of each bit is chosen uniformly and independently among the 2^t ancestors of the given individual (this is specific to uniform crossover). In other words, the distribution of the ancestors of the n bits of an individual is an independent sampling with replacement of n individuals among its 2^t ancestors.

We will use the fact that, sometimes, two bits come from the same ancestor to evaluate the deviation of p_t from p_∞. For this purpose, we will take as our p_0 the distribution on $\{0,1\}^n$ putting weight $1/2$ on the individual $111\ldots1$ and $1/2$ on $000\ldots0$. We will consider the law of the number of 1's in an individual under p_∞ and p_t, and find a difference.

Under p_∞, the law of the number of 1's is binomial with parameters n and $1/2$.

Under p_t, each bit of an individual comes from one of its ancestors at time 0. If the n ancestors of the n bits are all distinct, then these bits are picked uniformly and independently from p_0, in which case we find again a binomial distribution.

If, conversely, two bits of an individual come from the same ancestor at time 0, given our population p_0, these two bits will be equal. This leads to correlations which result in a quantifiable difference in the law of the number of 1's in an individual.

We will first evaluate the deviation obtained when exactly two bits have the same ancestor. We will then show that exactly two bits have the same ancestor with a large enough probability, and that the cases when more than one correlation occurs have a negligible weight when t is large. The first statement is the subject of the following lemma.

Lemma 6.
Let $n \geqslant 8$. Let μ_1 be the uniform probability measure on $\{0,1\}^n$. Let μ_2 be the measure on $\{0,1\}^n$ equal to $1/2^{n-1}$ at those points $x = (x_1,\ldots,x_n) \in \{0,1\}^n$ such that $x_1 = x_2$ and equal to 0 elsewhere. Then, the difference between the probabilities under μ_1 and μ_2 of the event "the number of 1's in a sample individual lies between $n/2 - \sqrt{n/8}$ and $n/2 + \sqrt{n/8}$" is larger than $1/2n$.

Proof of the lemma.
Under μ_1, without correlated bits, the law of the number of 1's is a binomial $\binom{n}{r}/2^n$.

Under μ_2, there is one pair of correlated bits, and the law of the number of 1's will rather be $\frac{1}{2}\binom{n-2}{r-2}/2^{n-2} + \frac{1}{2}\binom{n-2}{r}/2^{n-2}$ (respectively for the cases when the correlated pair is made up of two 1's or two 0's).

The latter is less than the former in a zone around $n/2$, and greater elsewhere. The difference between the two is $\left(\binom{n}{r} - 2\binom{n-2}{r-2} - 2\binom{n-2}{r}\right)/2^n$, which is, after a small calculation, equal to $\binom{n-2}{r-1}/2^n \left(\frac{n-(n-2r)^2}{2(n-r)r}\right)$. The second term is positive for $n/2-\sqrt{n}/2 \leqslant r \leqslant n/2+\sqrt{n}/2$, equal to $2/n$ at $r = n/2$; it is greater than $1/n$ for $|r - n/2| \leqslant \sqrt{n/8}$.

Then, the difference of the probabilities under μ_1 and μ_2 that the number of 1's falls between $n/2 - \sqrt{n/8}$ and $n/2 + \sqrt{n/8}$ is greater than

$$\frac{1}{n}\sum_{r=n/2-\sqrt{n/8}}^{n/2+\sqrt{n/8}} \frac{1}{2^n}\binom{n-2}{r-1}$$

Knowing that a binomial of parameter $1/2$ is almost a bell curve:

$$\sum_{r=n/2-\sqrt{n/8}}^{n/2+\sqrt{n/8}} \frac{1}{2^{n-2}}\binom{n-2}{r-1} \sim \frac{2}{\sqrt{\pi}}\int_{-1/\sqrt{2}}^{1/\sqrt{2}} e^{-x^2/2}\mathrm{d}x \geqslant 1/2$$

(and the first term is indeed greater than $1/2$ as soon as $n \geqslant 8$), we get that this expression is greater than $1/2n$, which proves the lemma. $\square$

Observe that, if the n bits of an individual at step t have distinct ancestors at step 0, the law of the number of 1's in these n bits is the same as under μ_1 in the lemma. If exactly two bits have the same ancestor, the law of the number of 1's will be the same as under μ_2 in the lemma.

We will now derive from this an evaluation of the distance between p_t and p_∞. Let A_0 be the event "all n bits have distinct ancestors at time 0", A_1 the event "exactly one pair of bits has a common ancestor", A_2 the remaining cases (more than one coincidence). Let also B be the event "the number of 1's falls between $n/2 - \sqrt{n/8}$ and $n/2 + \sqrt{n/8}$".

Then, according to the lemma:

$$\begin{aligned}
|p_\infty - p_t| &\geqslant |p_\infty(B) - p_t(B)| \\
&\geqslant |(p_\infty(B) - p_t(B|A_0))\, p_t(A_0) + (p_\infty(B) - p_t(B|A_1))\, p_t(A_1)| \\
&\quad - |p_\infty(B) - p_t(B|A_2)|\, p_t(A_2) \\
&\geqslant 0 + \frac{1}{2n} p_t(A_1) - p_t(A_2)
\end{aligned}$$

(Knowing A_0, the number of 1's under p_t is the same as under p_∞.)

Thus, the issue is to evaluate the probabilities that exactly two bits, or more than two bits, have a common ancestor. We know that these ancestors are sampled uniformly and independently from a set of size 2^t. We state the following lemma, which we will use again later.

Lemma 7.
If n (distinguishible) individuals are placed uniformly at random into k cells, the probability that exactly two elements are placed in the same cell is greater than $\frac{n^2}{4k}\left(1-\frac{n^2}{2k}\right)$.

Proof of the lemma.
By elementary combinatorics, this probability is $\frac{1}{k^n}\frac{n(n-1)}{2}\,k(k-1)\ldots(k-n+2)$, that is $\frac{n(n-1)}{2k}\,1\,(1-1/k)\ldots(1-(n-2)/k)$, which is greater than $\frac{n^2}{4k}\left(1-\frac{n^2}{2k}\right)$. □

Let's denote $k=2^t$. By Lemma 7, the probability that exactly two bits of an individual have the same ancestor at time 0 is more than $n^2/8k$ for k large enough. The case when more than two correlations would occur, that is, at least two pairs of bits with common ancestors, or at least three bits with the same ancestor, has a probability not greater than n^4/k^2, which is of greater order in $1/k$. Indeed, the probability that two pairs have common ancestors is at most $(n(n-1)/2)^2\,k(k-1)k^{n-4}/k^n = O(n^4/k^2)$, and the probability that three bits have the same ancestor is $(n(n-1)(n-2)/6)\,k^{n-2} = O(n^3/k^2)$.

We saw above that $|p_\infty - p_t| \geqslant 1/(2n)\,p_t(A_1) - p_t(A_2)$. If we take $k \geqslant 16n^3$ we ensure that the probability of A_2 is less than $n/32k$, in which case the expression at play is no less than $n/32k$. □

2.B Lower bound for finite populations

It is instructive to note that the difference between the laws q_∞ and p_∞ cannot be interpreted as an error due to the sampling with replacement in the k-tuple π_0 of the genes of an individual of q_∞, as opposed to sampling without replacement in p_∞. Indeed, if that were the case, the probability that two genes of an individual of π_∞ come from the same individual in π_0 would be exactly $1/k$, whereas we have just seen that it is actually $1/(k(1-r_\Pi+1/k))$, which is greater especially for large populations.

The above analysis shows that $|q_\infty - p_\infty| \leqslant \frac{n^2}{k\,(1-r_\Pi+1/k)}$. Let's prove a corresponding lower bound, which shows we cannot improve this result by much:

Theorem 8.
For all $n \geqslant 2$, for k large enough, there exists some initial population such that

$$|q_\infty - p_\infty| \geqslant \frac{n}{32k}$$

The k above which the proposition holds depends on n (otherwise, $n/32k$ could be more than 1). Since this is a negative asymptotic result, we will not worry too much about an explicit value for the k above which the proposition holds; a crude inspection of the proof reveals it holds at least for $k \geqslant 48(2n)^{n^2+2}/(1-r_\Pi)^{n^2}$. Of course this is probably a gross overestimate.

Proof.
As usual, we will consider an individual at time t, and look at the individuals at time 0 from which its n bits arise. We will have a close look at the distribution of these n individuals.

We will essentially work as in section 2.A: we will show that, with some probability of order n^2/k, exactly two bits have the same ancestor, which introduces a deviation of order $1/n$.

First, we will evaluate the probability that exactly two bits have a common ancestor at time 0. This probability is greater than the probability that exactly two bits have the same ancestor at time 0 and that, in addition, all bits are separated at time 1.

In the proof of theorem 4, we saw that the probability that some two bits of an individual of π_∞ have the same ancestor at time 1 is less than $\dfrac{n^2}{k(1-r_\Pi+1/k)}$. Thus, the probability that all of them are separated at time 1 is greater than $1-\dfrac{n^2}{k(1-r_\Pi+1/k)}$.

Now, if all bits are separated at time 1, their parents at time 0 are simply picked uniformly and independently among k. According to lemma 7, the probability that exactly two of them fall together is greater than $(n^2/4k)(1-n^2/2k)$.

Thus, the (unconditional) probability that at time 0, exactly two bits fall together is greater than $\dfrac{n^2}{4k}\left(1-\dfrac{n^2}{2k}\right)\left(1-\dfrac{n^2}{k(1-r_\Pi+1/k)}\right)$ which in turn is more than $\dfrac{n^2}{8k}$ as soon as k is large enough, say $k \geqslant 3n^2/(1-r_\Pi)$.

Under the assumption that there exist two bits with the same ancestor, we will find a deviation between the probabilities of some event under p_∞ and q_t. Of course, we will take as our p_0 the probability distribution on $\{0,1\}^n$ which puts weight 1/2 on $111\ldots1$ and 1/2 on $000\ldots0$. Then, we will be interested in the distribution of the number of 1's in an individual of generation t.

We will argue as in section 2.A. To do this, we must first establish that the case when exactly two bits of an individual at time t have the same ancestor at time 0 is predominant over the cases when there are more coincidences. This is the subject of the following lemma, which states that the distribution of the ancestors of the n bits of an individual has roughly the same asymptotics, when $k \to \infty$, as if these ancestors were sampled uniformly and independently among the k individuals of the initial population.

In particular, the cases when exactly two bits have the same ancestor will have a probability of order $1/k$, whereas those when more correlations occur will weigh for less than $1/k^2$. We measure the number of coincidences by the number of distinct individuals from which the n bits of our individual at time t come from. This lemma can be of independent interest.

Lemma 9.
There exist constants $C_{n,\Pi}$ and $C'_{n,\Pi}$ such that the probability that the n bits of an

individual at time $t = \infty$ *come from* m *distinct individuals from time* 0 *lies between* $\frac{C_{n,\Pi}}{k^{n-m}}$ *and* $\frac{C'_{n,\Pi}}{k^{n-m}}$, *for* k *large enough.*

(It is easy to see that it makes sense to speak about an individual from generation $t = \infty$: the process is Markovian on the space of k-individual populations. Often we will look at the process backwards, as if it started at $t = \infty$; this can easily be made rigorous by taking t large enough afterwards.)

Proof of the lemma.

Let's fix an individual from generation t, $t \approx \infty$ (i.e. t large enough). We have already seen that for $n = m$, the probability that all its bits have distinct ancestors at time 0 is greater than $1 - O(1/k)$, for large t. (The constants implied in $O()$ depend of course on n and Π.)

The idea is to consider the Markov chain made up of the positions (in the k-individual population) of the ancestors of the n bits of the given individual, at time $t - t'$ (a Markov chain in t'). We will split this Markov chain into classes, the class m being made up of those situations when the n bits are distributed over $m \leqslant n$ individuals at time $t - t'$. We will consider the communication probabilities between these classes, and study the weight of these classes in equilibrium when t' tends to infinity (relative to t, but we take a large t).

Let $m(t')$ be the number of distinct individuals which the n bits come from at time $t - t'$, and $s(m)$ the probability that $m(\infty) = m$. We intend to show that $s(m) = O(1/k^{n-m})$. We already know that for $m < n - 1$, $s(m) = O(1/k)$. In the following, the constants implied by O depend on n, m and r_Π; we only intend to study the asymptotic behavior in k.

Now, let's estimate the distribution of $m(t' + 1)$ for a given $m(t')$.

To go from generation $t-t'$ to generation $t-t'-1$, we consider the $m(t')$ individuals carrying the n bits. We decompose the process into two steps. In the first one, we consider the $m(t')$ blocks of bits, and we apply the mating operator Π to find $2m(t')$ "abstract parents" generating them. Among these $2m(t')$, only m', where $m(t') \leqslant m' \leqslant n$, carry some bits. In the second step, we paste back these m' abstract parents onto the population at time $t - t' - 1$, which is made up of k individuals. The pasting consists in choosing, for each of the m' abstract parents, which individual among the k it really is. These individuals are chosen independenlty and uniformly among k (there is some additional complication due to the fact that the two parents of one individual are distinct, in which case we choose among $k - 1$ rather than k, which does not affect the calculation much).

The probability that these m' parents are spread over $m'' \leqslant m'$ individuals of generation $t - t' - 1$ is, by elementary combinatorics, of order $C_{m'}/k^{m'-m''}$ for large k. Now, knowing $m(t')$, we know that $m' \geqslant m(t')$ and that, moreover, if $m(t') < n$, then $m' > m(t')$ with probability greater than $1 - r_\Pi$.

In other words, the first of our two steps cannot decrease $m(t')$, and increases it with probability greater than $1 - r_\Pi$ (if $m(t) < n$); the second one decreases the result with controlled probability, going from m' to m'' with probability $O(1/k^{m'-m''})$. All in all, $m(t' + 1) < m(t')$ with probability $O(1/k^{m(t')-m(t'+1)})$, $m(t' + 1) = m(t')$ with probability less than $r_\Pi + O(1/k)$, and $m(t'+1) > m(t')$ otherwise: generally, the number of blocks of bits increases, and it decreases only with probabilities controlled by powers of k.

Let's move to the proof proper. We work by backwards induction on m.

Suppose we have already proved that for all $m' \leqslant m$, we have $s(m') = O(1/k^{n-m})$, and that for $m \leqslant m' \leqslant n$ we have $s(m') = O(1/k^{n-m'})$. Now, the probability $s(1)$ that at time 0 ($t' \approx \infty$), all bits lie together, is such that $s(1) \leqslant r_\Pi\, s(1) + O(1/k)\, s(2) + O(1/k^2)\, s(3) + \ldots + O(1/k^{n-1})\, s(n)$ (in equilibrium). According to our induction hypothesis, and since $r_\Pi < 1$, this is $O(1/k^{m+1})$.

Similarly, $s(2) \leqslant s(1) + r_\Pi\, s(2) + O(1/k)\, s(3) + \ldots + O(1/k^{n-2})\, s(n)$, which is $O(1/k^{m+1})$ by our induction hypothesis, and since $r_\Pi < 1$.

Step by step, up to $m' = m-1$, we get that for $m' \leqslant m-1$, we have $s(m) = O(1/k^{n-m+1})$, which concludes our induction and ends the proof of the upper bound in the lemma (the constants in the notation O depend on everything except k).

In order to get the lower bound in the lemma, it is enough to observe that $s(n) = 1 - O(1/k)$ and to note that the transition coefficients $n \to m$ from the state $m(t') = n$ to $m(t'-1) = m$ are of order $1/k^{n-m}$. □

On one hand, we proved that exactly two bits have a common ancestor with probability greater than $n^2/8k$; on the other hand, the case when more than one pair of bits have a common ancestor has probability at most $O(1/k^2)$. It is then enough to take k large and apply lemma 6 to conclude. □

3 Coalescence and mean-time approximation of a population

The results stated above deal with extraction of one individual from the finite population π_t. One can wonder if the law of the whole k-tuple π_t is close to, for example, the law $p_\infty^{\otimes k}$ of an independent k-sample from p_∞. This is false due to the coalescence phenomenon.

The following is a classical result in the so-called Wright-Fisher model (see e.g. [7], [3], [2] or [4]).

Proposition 10.
For large k, for all $\varepsilon > 0$, for

$$t \geqslant 4k\,(\ln n - \ln \varepsilon + \ln 2)$$

then, with probability greater than $1-\varepsilon$, the k-tuple π_t is made up of k copies of the same individual.

The k above which the proposition holds is independent of n and ε.

Corollary 11.
Under the same assumptions, the distance $\left|\sigma_t - p_\infty^{\otimes k}\right|$ *is greater than*

$$1 - \varepsilon - \prod_{1 \leqslant i \leqslant n} \left(a_i^k + (1-a_i)^k\right)$$

where σ_t *is the law of the* k*-tuple* π_t*, which is a probability distribution on* $(\{0,1\}^n)^k$.

Proof.
Indeed, $\prod_{1 \leqslant i \leqslant n} \left(a_i^k + (1-a_i)^k\right)$ is the weight, under $p_\infty^{\otimes k}$, of k-tuples made up of identical individuals. □

However, even for k not too large, the coalescence time $4k \log n$ is much larger than the characteristic time of the convergence $q_t \to p_\infty$, which is of order $2 \log_{1/r_\Pi} n$. So hopefully, in the meantime, some number $m \leqslant k$ of individuals could be extracted from π_t, whose joint law would be close to $p_\infty^{\otimes m}$.

Indeed:

Theorem 12.
Let $m \leqslant k$*. Let* q_t^m *be the joint law in* $(\{0,1\}^n)^m$ *of the first* m *individuals of* π_t*. Then*

$$\left|q_t^m - p_\infty^{\otimes m}\right| \leqslant \frac{m^2 n^2}{k(1 - r_\Pi + 1/k)} + \frac{m^2 n}{k}\, t + mn^2\, r_\Pi^t$$

Of course, "the m first individuals" could be replaced by any m-tuple chosen in advance among π_t.

The first term corresponds to the intrinsic bias of the finite population, even for long times, as studied above. The second reflects coalescence. The third renders the convergence to p_∞.

Note that k must be of order $(mn)^2$ for a non-trivial estimate.

The optimum in t (tradeoff between coalescence and convergence to p_∞) is achieved for $t \approx \log_{1/r_\Pi} \dfrac{nk}{m}$ and is roughly $\dfrac{m^2 n^2}{k(1 - r_\Pi + 1/k)} + \dfrac{nm^2}{k} \log_{1/r_\Pi} \dfrac{nk}{m}$.

Using the same techniques as before (evaluating the number of 1's among the mn bits when two bits have the same ancestor), one may derive a lower bound, which matches the upper bound up to a factor of $1/mn$ (and constants), for large k and a given t.

Proof.
We will follow the ancestry of the mn bits of the first m individuals of π_t. If these mn bits come from distinct individuals of π_0 (which requires $k \geqslant mn$), then the resulting distribution will be $p_\infty^{\otimes m}$.

Let us consider two given bits among these mn. If they are two different bits from the same individual, nothing changes in regard to our previous analysis, and the probability that they are not separated at time 0 is less than $\dfrac{1}{k(1-r_\Pi+1/k)}+r_\Pi^t$.

If these two bits are located at different positions in two different individuals of π_t, then the Markov chain describing their separation is the same. However, initially, they are separated. Their probability of falling together at a given time begins at 0 and tends geometrically to $\dfrac{1}{k(1-r_\Pi+1/k)}$; it is always less than $\dfrac{2}{k(1-r_\Pi+1/k)}$.

However, the picture is quite different if we consider two bits located at the same position in two individuals of π_t: indeed, if, somewhere in the family tree, these two bits are gathered into one single individual, they are actually the *same* bit, inherited from that individual. Going back further in the tree, up to π_0, the bits can never again be separated.

Given these two bits, at each (backward) generation, their gathering occurs when they have the same parent, i.e. with probability $1/k$. The probability of their gathering in t backward steps is, thus, less than t/k (which is essentially tight for large k).

There are $mn(n-1)/2$ pairs of bits at different positions in a single individual; $m(m-1)n(n-1)/2$ pairs of bits at different positions in two different individuals; and $m(m-1)n/2$ pairs of bits located at the same position in two distinct individuals. Hence the result, by the same reasoning as in theorem 4. □

Acknowledgements

I would like to thank Pierre Pansu for introducing me to the subject of genetic algorithms, for weekly conversations and numerous comments. Many thanks to Raphaël Cerf for having organised a small workshop on genetic algorithms I attended, and for helpful conversations. Thanks to Claire Kenyon as well for her presenting the work of Y. Rabani, Y. Rabinovich and A. Sinclair and for detailed comments on the manuscript. Thanks to the referee for useful comments.

References

[1] S. Arora, Y. Rabani, U. Vazirani, *Simulating quadratic dynamical systems is PSpace-complete*, Proc 26th ACM Symp Theory of Computing (1994), p. 459–467.

[2] W. J. Ewens, *Mathematical Population Genetics*, Springer-Verlag, Berlin (1979).

[3] R.A. Fisher, *The genetical theory of natural selection*, Clarendon Press, Oxford (1930).

[4] S. Karlin, H.Taylor, *A second course in stochastic processes*, Academic Press, New York (1981).

[5] Y. Rabani, Y. Rabinovich and A. Sinclair, *A computational view of population genetics*, Random Structures and Algorithms **12** (1998), 4, p. 313–334.

[6] Y. Rabinovich, A. Sinclair, A. Widgerson, *Quadratic dynamical systems*, Proc 33rd IEEE Symp Foundations of Computer Science (1992), p. 304–313.

[7] S. Wright, *Evolution in Mendelian populations*, Genetics **16** (1931), p. 97–159.

Habilitation à diriger des recherches

Large-scale non-linear effects of fluctuations in relativistic gravitation

Ce texte est une synthèse de quelques-uns des résultats obtenus par Claire Chevalier, Fabrice Debbasch et moi-même dans le cadre d'une théorie statistique de la relativité générale, l'un des thèmes du projet ANR que j'ai dirigé en 2005–2008. Cette théorie a été initiée il y a quelques années par Fabrice Debbasch, de l'université Paris 6. Cette synthèse, dont les résultats sont exposés plus en détail dans les textes suivants, est à paraître dans la revue Nonlinear Analysis: Theory, Methods and Applications.

Large-scale non-linear effects of fluctuations in relativistic gravitation

Claire Chevalier, Fabrice Debbasch & Yann Ollivier

Abstract

The first fully non-linear mean field theory of relativistic gravitation has been developed in 2004. The theory makes the striking prediction that averaging or coarse graining a gravitational field changes the apparent matter content of space-time. A review of the general theory is presented, together with applications to black hole and cosmological space-times. The results strongly suggest that at least part of dark energy may be the net large scale effect of small scale fluctuations around a mean homogeneous isotropic cosmology.

1 Introduction

General relativity is a non-linear theory and, as such, small-scale phenomena may have a non-trivial average effect at large scales. Since, at the same time, astrophysical and cosmological observations have only finite space and time resolutions, it is in practice necessary [1] to develop a mean field theory of gravitation, i.e. an effective theory allowing a self-consistent description of the observed gravitational field at a given scale or resolution, accounting for the average net effects of small-scale phenomena not accessible within a given observational setup.

Developing such an effective theory has long been the subject of active research ([2, 3, 4, 5, 6, 7, 8]). The first general mean field theory for Einstein gravitation has been obtained four years ago ([9, 10]). The theory makes the striking prediction that averaging or coarse graining a gravitational field changes the apparent matter content of space-time. In particular, the net 'large scale' effect of the averaged upon, 'small scale' gravitational degrees of freedom is to contribute an 'apparent matter' at large scale, necessary to account for the coarse grained gravitational field. This matter may be charged if the gravitational field is coupled to an electromagnetic field.

This contribution is organized as follows. We first introduce the general mean field theory. Then we address perturbatively the important example of background gravitational waves around a simple homogeneous and isotropic, spatially flat dust universe; our results show, at least for this very simple model, that there is a frequency and amplitude range in which background waves, while being undetectable with current techniques, would generate an apparent large scale matter of energy density comparable to the energy density of the dust.

Finally we present coarse grainings of both the Schwarzschild and the extreme Reisner–Nordström (RN) black holes. In particular, the Schwarzschild black hole, which is a vacuum solution of the Einstein field equations, is shown to appear, after coarse-graining, as surrounded by an apparent matter whose equation of state strongly resembles the equation of state commonly postulated for cosmological dark energy. We also investigate thermodynamical aspects, highlighting the fact that the envisaged coarse graining transforms the extreme RN black hole, which has a vanishing temperature, into a black hole of non-vanishing temperature.

2 A mean field theory for general relativity

Let $\mathcal{M}$ be a fixed manifold and let Ω be an arbitrary probability space. Let $g(\omega)$ be an ω-dependent Lorentzian metric defined on $\mathcal{M}$; let also $A(\omega)$ be an ω-dependent electromagnetic 4-potential, with associated current $j(\omega)$. Each triplet $\mathcal{S}(\omega) = (\mathcal{M}, g(\omega), A(\omega))$ represents a physical space-time depending on the random parameter $\omega \in \Omega$. For example, $g(\omega)$ may represent a gravitational wave of random phase and wave vector around a given reference space-time.

With each space-time $\mathcal{S}(\omega)$ are associated the Einstein tensor $E(\omega)$ of the metric $g(\omega)$, and a stress-energy tensor $\mathcal{T}(\omega)$ satisfying the Einstein equation

$$E(\omega) = 8\pi\mathcal{T}(\omega). \tag{1}$$

We decompose

$$\mathcal{T}(\omega) = \mathcal{T}_{(A(\omega),g(\omega))} + \mathcal{T}_m(\omega), \tag{2}$$

where $\mathcal{T}_{(A(\omega),g(\omega))}$ represents the electromagnetic stress-energy tensor generated by $A(\omega)$ in $g(\omega)$, and $\mathcal{T}_m(\omega)$ represents the stress-energy of other matter fields.

It has been shown in ([9]) that such a collection of space-times can be used to define a single, mean or coarse grained space-time $(\mathcal{M}, \bar{g}, \bar{A})$ representing the average, "macroscopic" behavior of these random spaces-times. The metric $\bar{g}$ and the potential $\bar{A}$ are the respective averages of the metrics $g(\omega)$ and of the potentials $A(\omega)$ over ω; thus, for all points P of $\mathcal{M}$,

$$\bar{g}(P) = \langle g(P, \omega)\rangle \tag{3}$$

and

$$\bar{A}(P) = \langle A(P, \omega)\rangle, \tag{4}$$

where the brackets on the right-hand side indicate an average over the probability space Ω. If we think of $g(\omega)$ as a reference metric perturbed by small random contributions then $\bar{g}$ will represent the average metric, where the fluctuations have been smoothed out but with the same macroscopic behavior.

The metric $\bar{g}$ defines an Einstein tensor $\bar{\mathcal{E}}$ for the coarse-grained space-time. However, since the expression for the Einstein tensor as a function of the metric is non-linear, this Einstein tensor and the average stress-energy tensor will, in general, *not* be related by the Einstein equation:

$$\bar{\mathcal{E}} \neq 8\pi\langle\mathcal{T}(\omega)\rangle$$

so that physical measurements attempting to relate the coarse-grained space-time to its average matter content would yield to a violation of the Einstein equation. This would not happen in a Newtonian setting since, then, the relation between field and matter is linear and thus is unchanged under averaging.

To enforce validity of the Einstein equation for the coarse-grained space-time, it is thus necessary to introduce a new term

$$T^{\mathrm{app}}{}^{\alpha}_{\beta} = \bar{\mathcal{E}}^{\alpha}_{\beta}/8\pi - \langle (\mathcal{T}_m)^{\alpha}_{\beta}(\omega)\rangle - \mathcal{T}_{(\bar{A},\bar{g})}{}^{\alpha}_{\beta}, \tag{5}$$

so that the stress-energy tensor of the coarse-grained space-time can be described as the sum of the stress-energy tensor of the average quadripotential $\bar{A}$, of the average stress-energy tensor $\langle(\mathcal{T}_m)(\omega)\rangle$ appearing in the averaged spaces-times, and of this new term T^{app}. This generally non-vanishing tensor field can be interpreted as the stress-energy tensor of an 'apparent matter' in the coarse-grained space-time. This apparent matter describes the cumulative non-linear effects of the averaged-out small scale fluctuations of the gravitational and electromagnetic fields on the large scale behaviour of the coarse-grained gravitational field.

In particular, even the vanishing of $\mathcal{T}(\omega)$ for all ω does not necessarily imply the vanishing of $\bar{\mathcal{T}}$. The mean stress-energy tensor $\bar{\mathcal{T}}$ can therefore be non-vanishing in regions where the unaveraged stress-energy tensor actually vanishes.

The Maxwell equation relating the electromagnetic potential to the electromagnetic current also couples the electromagnetic field and the gravitational field non-linearly; the mean current $\bar{j}$ associated with $\bar{A}$ in $\bar{g}$ does not therefore coincide with the average $\langle j(\omega)\rangle$. In particular, a region of space-time where $j(\omega)$ vanishes for all ω is generally endowed with a non-vanishing mean current $\bar{j}$.

Let us finally mention that the averaging scheme just presented is the only one which ensures that motion in the mean field can actually be interpreted, at least locally, as the average of 'real' unaveraged motions. This important point is fully developed in ([10]).

3 Waves around a homogeneous isotropic simple cosmology

The averaging procedure above has been applied to background gravitational waves [11] propagating around a homogeneous isotropic, spatially flat dust universe. The main conclusion is that the large-scale effect of these gravitational waves is close to that of a matter field with positive energy and pressure, whose order of magnitude is roughly $n^2\varepsilon^2$ where n is the relative frequency of the waves and ε their relative amplitude. In particular, in some regimes this energy would be comparable to that of the dust, even for some currently undetectable gravitational waves.

The reference metric and stress-energy tensor are the flat Friedman–Lemaître–Robertson–Walker (FLRW) universe, which, in conformal coordinates, reads:

$$g^{\mathrm{ref}} = a(\eta)^2(d\eta^2 - dx^2 - dy^2 - dz^2) \qquad T^0_0 = \rho(\eta) \quad T^0_i = T^j_i = 0 \tag{6}$$

where a is the so-called expansion factor and ρ is the energy density. The Einstein equation delivers $a(\eta) = C\eta^2$ and $8\pi\rho(\eta) = 3\dot{a}^2/a^4 = 12/C^2\eta^6$, with C an arbitrary (positive) constant. Proper time is $\tau = C\eta^3/3$ and the Hubble 'constant' is $H = \frac{1}{a}\frac{da}{d\tau} = \frac{\dot{a}}{a^2} = \frac{2C}{\eta^3}$.

We will assume that the averaging scale is much larger than the wavelength of the gravitational waves; this means that we can represent the waves as a statistical ensemble of Fourier series with random phase $\omega \in [0; 2\pi]$.

So let us consider a gravitational wave propagating around the homogeneous and isotropic FLRW background. By isotropy we can assume that the wave propagates in the direction x. Such a gravitational wave is represented at first order by the metric perturbation

$$g^{(1)}{}_{22} = -\varepsilon(\eta)\, a(\eta)^2 e^{iq(x-\eta)}(1 - i/q\eta) \qquad g^{(1)}{}_{33} = -g^{(1)}{}_{22} \tag{7}$$

for the first polarization (the other polarization yields identical results and thus will not be discussed). Here the constant q is the wave number in conformal coordinates, and the relative amplitude of the wave is given by $\varepsilon(\eta) = \varepsilon_0/\eta^2$ for some constant ε_0. The number of oscillations (periods) in that part of the universe accessible to an observer situated at time η is $n_{\text{osc}} = q\eta$.

Using the real part of the above, i.e. $-\varepsilon(\eta)\, a(\eta)^2(\cos(q(x-\eta)) + \frac{1}{q\eta}\sin(q(x-\eta)))$, we can compute the apparent matter associated with this gravitational wave and compare it to the energy density $\rho(\eta)$ of the dust. Using (5) we get, at second order in ε and for large n_{osc}:

$$T^{\text{app}}{}^0_0 = \varepsilon(\eta)^2 n_{\text{osc}}^2\, \frac{1}{48}\rho(\eta) \tag{8}$$

$$T^{\text{app}}{}^1_1 = -\varepsilon(\eta)^2 n_{\text{osc}}^2\, \frac{1}{48}\rho(\eta) \tag{9}$$

$$T^{\text{app}}{}^1_0 = \varepsilon(\eta)^2 n_{\text{osc}}^2\, \frac{1}{48}\rho(\eta). \tag{10}$$

All other components are 0 and all these relations hold up to $O(\varepsilon^3 n_{\text{osc}}^2 + \varepsilon^2)$.

Consider now a superposition of statistically independent gravitational waves of type (7). Suppose that these waves share a common frequency and amplitude, but propagate along spatial directions which are distributed uniformly over the unit sphere. The stress-energy tensor associated with the superposition of these waves can be easily deduced from the above and its non-vanishing components read:

$$T^{\text{app}}{}^0_0 = \varepsilon(\eta)^2 n_{\text{osc}}^2\, \frac{1}{48}\rho(\eta) \tag{11}$$

$$T^{\text{app}}{}^1_1 = T^{\text{app}}{}^2_2 = T^{\text{app}}{}^3_3 = -\varepsilon(\eta)^2 n_{\text{osc}}^2\, \frac{1}{144}\rho(\eta) \tag{12}$$

with all other components zero, up to the same order as before. The above expressions show that a background of high frequencies (*i.e.* $n_{\text{osc}} \gg 1$) gravitational waves behaves like radiation with positive pressure equal to a third of its energy density. The energy density is (at this order) quadratic in the frequency and amplitude.

Thus, a background wave of relative amplitude $\varepsilon \approx 10^{-5}$ and oscillation number $n_{\mathrm{osc}} \approx 10^5$ would generate an effective large-scale stress-energy in the universe comparable to the energy density of dust present in this model. Such a wave would correspond today to a physical frequency of order $10^{-12}\,\mathrm{Hz}$ and would elude direct observation [12].

4 Coarse graining of black hole space-times

Both the Schwarzschild and the extreme Reisner–Nordström space-times of total mass M have been coarse grained using the above procedure [13, 14, 15, 16]. For the Schwarzschild (resp. extreme RN) black hole, the metric $g(\omega)$ is the Schwarzschild (resp. extreme RN) metric spatially translated by ω (resp. $i\omega$) in spatial Kerr-Schild coordinates [17], with ω distributed uniformly in the 3-ball $\mathcal{B}_a$ of radius $a > 0$. In both cases, exact expressions have been found for the mean metric $\bar{g}$ for all points with radial Kerr-Schild coordinate r greater than the coarse graining parameter a.

Both averaged space-times describe black holes with the following properties. The total mass of the space-times, as well as the total charge of the extreme black hole are preserved by the averaging. But energy and mass are spatially redistributed: in particular the averaged Schwarzschild black hole is surrounded by an apparent matter with energy density ε equal to the opposite of the radial pressure p_r and the scalar curvature induced by this apparent matter is strictly negative. The similarities with dark energy are striking.

The temperature of the black holes are also modified by the averaging. Quite remarkably, the extreme black hole of vanishing temperature is modified into a finite temperature black hole. Indeed, the temperature of the black hole obtained by averaging the extreme RN black hole reads, at first order in the coarse graining parameter a:

$$\Theta(a, M) \simeq \frac{a}{2\sqrt{5}\pi\, M^2}. \qquad (13)$$

Thus, at least some classical black holes of finite temperature can be understood as statistical superposition of other *purely classical* (as opposed to quantum) gravitational fields.

5 Conclusion

We have reviewed the new mean field theory of relativistic gravitation and discussed some applications to black hole physics and cosmology. There are two main conclusions. The first one concerns black hole thermodynamics. We have proved that it is possible to build at least some finite temperature black holes as statistical ensembles of *classical* vanishing temperature extreme black holes. This result is striking because black hole thermodynamics was until now understood only by building black holes as statistical ensembles of quantum objects.

The second conclusion concerns cosmology. We have proved that small scale background gravitational waves propagating around an homogeneous and isotropic universe can contribute significantly to the large scale energy density. For instance, waves with a present relative amplitude of approximately 10^{-5} and a dimensionless comoving wave (oscillation) number of 10^5 would elude current observation and would be sufficient to close the universe.

References

[1] G.F.R. Ellis and H. van Elst. Cosmological Models. In M. Lachièze-Rey, editor, *Theoretical and observational cosmology : proceedings of the NATO Advanced Study Institute on Theoretical and Observational Cosmology, Cargese, France*, NATO science series. Series C, Mathematical and Physical Sciences. Kluwer Academic, Boston, 1999.

[2] T. Futamase. A new description for a realistic inhomogeneous universe in general relativity. *Prog. Theor. Phys.*, 86:389–99, 1991.

[3] M. Kasai. Construction of Inhomogeneous universes which are Friedmann-Lemaitre-Robertson-Walker on average. *Phys. Rev. D*, 69:2330–2, 1992.

[4] T. Futamase. General relativistic description of a realistic inhomogeneous universe. *Prog. Theor. Phys.*, 89:581–97, 1993.

[5] T. Futamase. Averaging of a locally inhomogeneous realistic universe. *Phys. Rev. D*, 53:681–9, 1996.

[6] R. Zalaletdinov. Averaging problem in general relativity, macroscopic gravity and using Einstein's equation in cosmology. *Bull. Astron. Soc. India*, 25:401–416, 1997.

[7] T. Buchert. On average properties of inhomogeneous fluids in general relativity: dust cosmologies. *Gen. Rel. Grav.*, 32:105–26, 2000.

[8] T. Buchert. On average properties of inhomogeneous fluids in general relativity: perfect fluid cosmologies. *Gen. Rel. Grav.*, 33:1381–405, 2001.

[9] F. Debbasch. What is a mean gravitational field? *Eur. Phys. J. B*, 37(2):257–270, 2004.

[10] F. Debbasch. Mean field theory and geodesics in general relativity. *Eur. Phys. J. B*, 43(1):143–154, 2005.

[11] R.A. Isaacson. Gravitational radiation in the limit of high frequency. II. Nonlinear terms and the effective stress tensor. *Phys. Rev.*, 166(5):1272, 1968.

[12] W. Zhao and Y. Zhang. Relic gravitational waves and their detection. *Phys. Rev. D*, 74:043503, 2006.

[13] F. Debbasch and Y. Ollivier. Observing a Schwarzschild black-hole with finite precision. *Astron. Astrophys.*, 433(2):397–404, 2005.

[14] C. Chevalier, M. Bustamante, and F. Debbasch. Thermal statistical ensembles of black holes. *Phys. A*, 376:293–307, 2007.

[15] C. Chevalier and F. Debbasch. Is matter an emergent property of space-time? *Submitted for publication to the Proceedings of the International Conference on Complex Systems, Boston, 2007.*

[16] C. Chevalier and F. Debbasch. Thermal statistical ensembles of classical extreme black holes. *Submitted to Physica A*.

[17] S. Chandrasekhar. *The Mathematical Theory of Black-Holes.* Oxford University Press, Oxford, 1992.

Multiscale cosmological dynamics

Ce texte présente en détail certains résultats annoncés dans le texte précédent, obtenus avec Claire Chevalier, Fabrice Debbasch et moi-même dans le cadre du projet ANR que j'ai dirigé en 2005–2008. Il est actuellement soumis pour publication.

Yann Ollivier, École normale supérieure de Lyon, 2009

Multiscale cosmological dynamics

Claire Chevalier, Fabrice Debbasch & Yann Ollivier

Abstract

The recently developed mean field theory of relativistic gravitation predicts the emergence of an "apparent matter" field at large scales describing the net effect of small-scale fluctuations on the large-scale dynamics of the universe. It is found that this so-called back reaction effect is much stronger for gravitational waves than for matter density fluctuations. At large scales, gravitational waves behave like radiation and, for them, the perturbative effect scales as the squared relative amplitude times squared frequency. In particular, a bath of gravitational waves of relative amplitude 10^{-5} and frequency 10^{-12} Hz would not be directly detectable by today's technology but would generate an effective large-scale radiation of amplitude comparable to the unperturbed matter density of the universe.

1 Introduction

Multiscale systems are characterized by intricate dynamics which couple several different time or space scales. Nevertheless, in many instances it is possible to obtain an effective dynamics governing the evolution of a multiscale system on a larger scale by averaging the full exact dynamics on smaller scales. Examples range from economics [1] to biophysics [2] and include turbulence [3] and quantum field theory at both vanishing and finite temperature [4].

This article deals with relativistic gravitating systems. These are described, at the classical level, by general relativity [5], and are multiscale because Einstein's theory is strongly non-linear. The largest gravitating system is the universe and its large-scale description is the traditional object of cosmology. It is now well established [6] that the universe is, on large scale, expanding in an homogeneous and isotropic manner. Several authors [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17] have recently argued that small-scale fluctuations around this large-scale expansion may, by non-linearity, contribute substantially to the large-scale energy repartition generating the expansion. This article investigates the importance of this so-called "back reaction" effect on dust universes perturbed by background gravitational waves and small matter density fluctuations. Our main conclusion is that matter density perturbations produce a negligible back reaction but that, on the other hand, background gravitational waves may generate a large-scale energy density comparable to the energy density of dust. Implications of these results for physical cosmology, including the dark energy problem, are also discussed.

2 Mean field theory

2.1 Notation

In this article, the metric has signature $(+,-,-,-)$. We shall use mixed components T_μ^ν of the stress-energy tensor; with this signature, for a perfect fluid at rest with density ρ and pressure p we have $T_0^0 = \rho$ and $T_i^i = -p$.

2.2 General framework

Averaging classical gravitational fields necessitates a mean field theory of general relativity. Such a theory has been introduced in [7, 8]; we include a brief overview here for completeness. Some applications to black hole physics are presented in [18, 19, 20].

Let $\mathcal{M}$ be a fixed manifold, let Ω be an arbitrary probability space and let $g(\omega)$ be a Lorentzian metric on $\mathcal{M}$ depending on the random parameter ω. Each pair $\mathcal{S}(\omega) = (\mathcal{M}, g(\omega))$ represents a physical space-time depending on the random parameter $\omega \in \Omega$. For example, $g(\omega)$ may represent a gravitational wave of random phase and wave vector around a given reference space-time.

With each space-time $\mathcal{S}(\omega)$ are associated the Einstein tensor $G(\omega)$ of the metric $g(\omega)$, and a stress-energy tensor $\mathcal{T}(\omega)$, satisfying the Einstein equation

$$G(\omega) = 8\pi\mathcal{T}(\omega). \tag{1}$$

As shown in [7], such a collection of space-times can be used to define a single mean space-time $(\mathcal{M}, \bar{g})$ representing the average, "macroscopic" behavior of these random space-times. The metric $\bar{g}$ is the average of the metrics $g(\omega)$; thus, at every point P of $\mathcal{M}$,

$$\bar{g}(P) = \langle g(P, \omega) \rangle . \tag{2}$$

where the brackets on the right-hand side denote an average over the random parameter ω. If we think of $g(\omega)$ as a reference, "macroscopic" metric perturbed by small random contributions then $\bar{g}$ will represent the average metric, where the fluctuations have been smoothed out but with the same macroscopic behavior.

The metric $\bar{g}$ defines an Einstein tensor $\bar{G}$ for the mean space-time. However, since the expression for the Einstein tensor as a function of the metric is non-linear, this Einstein tensor and the average stress-energy tensor will, in general, *not* be related by the Einstein equation:

$$\bar{G} \neq 8\pi \langle \mathcal{T}(\omega) \rangle ,$$

so that physical measurements attempting to relate the mean space-time to its average matter content would yield to a violation of the Einstein equation. This would not happen in a Newtonian setting since, then, the relation between field and matter is linear and thus is unchanged under averaging.

To enforce validity of the Einstein equation for the mean space-time, it is thus necessary to introduce a new term

$$\mathcal{T}^{\mathrm{app}}{}_\mu^\nu = \bar{G}_\mu^\nu / 8\pi - \left\langle \mathcal{T}_\mu^\nu(\omega) \right\rangle , \tag{3}$$

so that the stress-energy tensor of the mean space-time can be described as the sum of the average stress-energy tensor $\langle \mathcal{T}(\omega)\rangle$ appearing in the averaged space-times, and of this new term $\mathcal{T}^{\mathrm{app}}$. This generally non-vanishing tensor field can be interpreted as the stress-energy tensor of an "apparent matter" in the mean space-time. Apparent matter describes the cumulative non-linear effects of the averaged-out small-scale fluctuations on the large-scale behaviour of the mean gravitational field.

In particular, even the vanishing of $\mathcal{T}(\omega)$ for all ω does not necessarily imply the vanishing of $\bar{\mathcal{T}}$. The mean stress-energy tensor $\bar{\mathcal{T}}$ can therefore be non-vanishing in regions where the unaveraged stress-energy tensor actually vanishes. We will see that this happens, for instance, with gravitational waves.

Note that the averaging scheme just presented is the only one which ensures that motion in the mean field can be interpreted, at least locally, as the average of "real" unaveraged motions. This important point is fully developed in [7, 8], together with an extension including non-quantum electrodynamics.

2.3 Small amplitude fluctuations

We now investigate the case when the metrics $g_{\mu\nu}(\omega)$ are all close to a reference metric ${g^{\mathrm{ref}}}_{\mu\nu}$. More precisely, we assume that there is a small parameter ε such that, for any value of the random parameter ω,

$$g_{\mu\nu}(\omega) = {g^{\mathrm{ref}}}_{\mu\nu} + \varepsilon {g^{(1)}}_{\mu\nu}(\omega) + \varepsilon^2 {g^{(2)}}_{\mu\nu}(\omega) + O(\varepsilon^3) \tag{4}$$

and we will expand the theory above at second order in ε.

Of course, any arbitrary choice of $g^{(1)}$ and $g^{(2)}$ will define a solution of the Einstein equation by setting the value of the stress-energy tensor to $\mathcal{T} = G/8\pi$, but these solutions are physically relevant only if the associated stress-energy tensor has a physical interpretation. In the sequel we will focus on choices of $g^{(1)}$ and $g^{(2)}$ arising from physically interesting stress-energy tensors, such as gravitational waves or fluctuations of the density of matter.

We now derive a perturbative expression for ${\mathcal{T}^{\mathrm{app}}}_{\mu}^{\nu}$. Denote by $\mathcal{D}G$ and $\mathcal{D}^2G$, respectively, the functional derivative and the functional Hessian of the Einstein tensor $G_\mu^\nu(g^{\mathrm{ref}})$ with respect to variations of the metric g^{ref}. So by definition we have the expansion

$$\begin{aligned} G_\mu^\nu(g(\omega)) = G_\mu^\nu(g^{\mathrm{ref}}) + \varepsilon(\mathcal{D}G_\mu^\nu)(g^{(1)}(\omega)) + \varepsilon^2(\mathcal{D}G_\mu^\nu)(g^{(2)}(\omega)) \\ + \frac{\varepsilon^2}{2}(\mathcal{D}^2G_\mu^\nu)(g^{(1)}(\omega), g^{(1)}(\omega)) + O(\varepsilon^3) \end{aligned} \tag{5}$$

which yields

$$\begin{aligned} 8\pi \left\langle T_\mu^\nu(\omega)\right\rangle = \left\langle G_\mu^\nu(g(\omega))\right\rangle = G_\mu^\nu(g^{\mathrm{ref}}) + \varepsilon \left\langle(\mathcal{D}G_\mu^\nu)(g^{(1)}(\omega))\right\rangle + \varepsilon^2 \left\langle(\mathcal{D}G_\mu^\nu)(g^{(2)}(\omega))\right\rangle \\ + \frac{\varepsilon^2}{2} \left\langle(\mathcal{D}^2G_\mu^\nu)(g^{(1)}(\omega), g^{(1)}(\omega))\right\rangle + O(\varepsilon^3) \end{aligned} \tag{6}$$

It is important to note here that $\mathcal{D}G$, being a functional derivative, is by definition a linear operator in its arguments $g^{(1)}$ or $g^{(2)}$. One thus has

$$\left\langle (\mathcal{D}G^\nu_\mu)(g^{(1)}(\omega))\right\rangle = (\mathcal{D}G^\nu_\mu)\left(\left\langle g^{(1)}(\omega)\right\rangle\right) \tag{7}$$

and likewise for $g^{(2)}$. But this is not true of the Hessian $\mathcal{D}^2G$, which is a quadratic (as opposed to linear) operator.

Meanwhile, the mean metric $\bar{g}$ is given by

$$\bar{g}_{\mu\nu} = g^{\mathrm{ref}}{}_{\mu\nu} + \varepsilon\left\langle g^{(1)}{}_{\mu\nu}(\omega)\right\rangle + \varepsilon^2\left\langle g^{(2)}{}_{\mu\nu}(\omega)\right\rangle + O(\varepsilon^3) \tag{8}$$

so that the associated Einstein tensor is

$$\begin{aligned}\bar{G}^\nu_\mu = G^\nu_\mu(\bar{g}) = G^\nu_\mu(g^{\mathrm{ref}}) + \varepsilon(\mathcal{D}G^\nu_\mu)\left(\left\langle g^{(1)}(\omega)\right\rangle\right) + \varepsilon^2(\mathcal{D}G^\nu_\mu)\left(\left\langle g^{(2)}(\omega)\right\rangle\right)\\ + \frac{\varepsilon^2}{2}(\mathcal{D}^2G^\nu_\mu)\left(\left\langle g^{(1)}(\omega)\right\rangle, \left\langle g^{(1)}(\omega)\right\rangle\right) + O(\varepsilon^3)\end{aligned} \tag{9}$$

From these results, by comparing $\bar{G}^\nu_\mu$ to $\left\langle G^\nu_\mu(g(\omega))\right\rangle$ we can directly compute the apparent stress-energy tensor:

$$\mathcal{T}^{\mathrm{app}\,\nu}{}_{\mu} = \frac{\varepsilon^2}{16\pi}\left((\mathcal{D}^2G^\nu_\mu)\left(\left\langle g^{(1)}(\omega)\right\rangle, \left\langle g^{(1)}(\omega)\right\rangle\right) - \left\langle(\mathcal{D}^2G^\nu_\mu)(g^{(1)}(\omega), g^{(1)}(\omega))\right\rangle\right) + O(\varepsilon^3) \tag{10}$$

which is generally non-zero due to the quadratic nature of $\mathcal{D}^2G$.

It is to be noted that the effect is at second order in ε, which was to be expected since at first order, gravitation is by definition linear. What is more interesting is that $g^{(2)}$ does not appear in the result. This reflects the fact that non-linearities acting on the second-order term $g^{(2)}$ will only produce higher-order terms.

It often makes sense to define the fluctuations in terms of the sources rather than the metric, i.e. to prescribe physically meaningful fluctuations $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ of the stress-energy tensor and to look for $g^{(1)}$ and $g^{(2)}$ solving the Einstein equation. That $g^{(2)}$ vanishes from the result means that, to compute the effect, it is actually enough to solve the linearized Einstein equation around g^{ref}.

A case of particular interest is when the fluctuations are "centered" i.e. when the average of the fluctuations is zero at first order:

$$\left\langle g^{(1)}(\omega)\right\rangle = 0 \tag{11}$$

in which case we simply get

$$\mathcal{T}^{\mathrm{app}\,\nu}{}_{\mu} = -\frac{\varepsilon^2}{16\pi}\left\langle(\mathcal{D}^2G^\nu_\mu)(g^{(1)}(\omega), g^{(1)}(\omega))\right\rangle \tag{12}$$

at this order in ε.

3 Fluctuations around dust cosmologies

3.1 Basics

We now apply the above to the case of either gravitational waves or density fluctuations around a homogeneous and isotropic, spatially flat dust universe (flat Friedmann–Lemaître–Robertson–Walker metric). The reference metric and stress-energy tensor of such a space-time are, in conformal coordinates [21]:

$$g^{\text{ref}} = a(\eta)^2(d\eta^2 - dx^2 - dy^2 - dz^2) \qquad T^0_0 = \rho(\eta) \quad T^0_i = T^j_i = 0 \tag{13}$$

where a is the so-called expansion factor and ρ is the energy density. The Einstein equation delivers $a(\eta) = C\eta^2$ and $8\pi\rho(\eta) = 3\dot{a}^2/a^4 = 12/C^2\eta^6$, where C is an arbitrary (positive) constant. Proper time is $\tau = C\eta^3/3$ and the Hubble "constant" is $H = \dfrac{1}{a}\dfrac{da}{d\tau} = \dfrac{\dot{a}}{a^2} = \dfrac{2C}{\eta^3}$.

The perturbations $g^{(1)}$ considered in this article will be of two types: gravitational waves and matter density fluctuations. They will be written as sums of spatial Fourier modes (this makes sense since g^{ref} is spatially flat). Each term in such a series is of the form $F(\eta)\exp\left(i(\mathbf{q}.\mathbf{r} + \omega_{\mathbf{q}})\right)$ where $F(\eta)$ is some tensor, $\mathbf{q}$ is a three-dimensional wave vector, and $\omega_{\mathbf{q}}$ is a phase associated with mode $\mathbf{q}$. Averaging a given mode $\mathbf{q}$ on spatial scales much larger than the wave-length $1/\left|\mathbf{q}\right|$ is equivalent to averaging this mode over the phase $\omega_{\mathbf{q}} \in [0; 2\pi]$. We therefore choose the set of all phases $(\omega_{\mathbf{q}})$ as our random parameter, and perform all averagings over these phases. For a superposition of statistically independent modes, the averaging can be performed independently for each $\omega_{\mathbf{q}}$.

3.2 Gravitational waves

Consider a single gravitational wave propagating along the above background [22]. This wave admits two polarizations [21]; since the background is isotropic, there is no loss of generality in assuming the wave propagates along, say the x-axis. The first-order metric perturbation then reads, for the first polarization:

$${g^{(1)}}_{22}(\omega) = -a(\eta)^2 e^{i(q(x-\eta)+\omega)}(1 - i/q\eta)/\eta^2 \qquad {g^{(1)}}_{33} = -{g^{(1)}}_{22} \tag{14}$$

with the other components equal to 0. Here q is the wave number in conformal coordinates, and $\omega \in [0; 2\pi]$.

The statistical averaging corresponds to a uniform averaging over $\omega \in [0; 2\pi]$; this models situations in which the system is observed at a resolution much larger than the perturbation wavelength $1/q$.

The quantity $n_{\text{osc}} = q\eta$ measures the typical number of oscillations (periods) in that part of the universe accessible to an observer situated at time η. The relative amplitude of the perturbation $\varepsilon g^{(1)}$ at time η, compared to g^{ref}, is $\tilde{\varepsilon}(\eta) = \varepsilon/\eta^2$. We will express the results in terms of those quantities.

The stress-energy tensor of apparent matter is then given by (12). In practice the Hessian term $\mathcal{D}^2 G^\nu_\mu$ in (12) is readily obtained as the ε^2 term in a Taylor expansion of the Einstein tensor of the metric $g^{\mathrm{ref}} + \varepsilon g^{(1)}$, which can be computed using any symbolic computation software.

Using the real part of the metric above, i.e. $-a(\eta)^2(\cos(q(x-\eta)+\omega)+\frac{1}{q\eta}\sin(q(x-\eta)+\omega))/\eta^2$, we get, at second order in ε :

$$\mathcal{T}^{\mathrm{app}}{}^0_0 = \tilde{\varepsilon}(\eta)^2 n_{\mathrm{osc}}^2\, \frac{1-14/n_{\mathrm{osc}}^2 - 39/2n_{\mathrm{osc}}^4}{48}\, \rho(\eta) \tag{15}$$

$$\mathcal{T}^{\mathrm{app}}{}^1_1 = -\tilde{\varepsilon}(\eta)^2 n_{\mathrm{osc}}^2\, \frac{1-2/n_{\mathrm{osc}}^2 - 27/2n_{\mathrm{osc}}^4}{48}\, \rho(\eta) \tag{16}$$

$$\mathcal{T}^{\mathrm{app}}{}^2_2 = \mathcal{T}^{\mathrm{app}}{}^3_3 = \tilde{\varepsilon}(\eta)^2 n_{\mathrm{osc}}^2\, \frac{1/n_{\mathrm{osc}}^2 + 9/2n_{\mathrm{osc}}^4}{48}\, \rho(\eta) \tag{17}$$

$$\mathcal{T}^{\mathrm{app}}{}^1_0 = \tilde{\varepsilon}(\eta)^2 n_{\mathrm{osc}}^2\, \frac{1}{48}\, \rho(\eta). \tag{18}$$

All other components are 0 and all these relations hold up to $O(\varepsilon^3 n_{\mathrm{osc}}^2 + \varepsilon^2)$.

If instead of a single wave, we consider a superposition of statistically independent gravitational waves sharing a common frequency and amplitude, but propagating along random spatial directions, we get a spatially isotropic version of (15–18), namely

$$\mathcal{T}^{\mathrm{app}}{}^0_0 = \tilde{\varepsilon}(\eta)^2 n_{\mathrm{osc}}^2\, \frac{1-14/n_{\mathrm{osc}}^2 - 39/2n_{\mathrm{osc}}^4}{48}\, \rho(\eta) \tag{19}$$

$$\mathcal{T}^{\mathrm{app}}{}^1_1 = \mathcal{T}^{\mathrm{app}}{}^2_2 = \mathcal{T}^{\mathrm{app}}{}^3_3 = -\tilde{\varepsilon}(\eta)^2 n_{\mathrm{osc}}^2\, \frac{1/3-4/3n_{\mathrm{osc}}^2 - 45/6n_{\mathrm{osc}}^4}{48}\, \rho(\eta) \tag{20}$$

These expressions show that a background of gravitational waves of high frequency ($n_{\mathrm{osc}} \gg 1$) behaves like an ordinary stress-energy tensor for radiation, with positive pressure equal to a third of its energy density (at this order in $\tilde{\varepsilon}(\eta)^2 n_{\mathrm{osc}}^2$).

The first-order metric perturbation for the other polarization is given by

$$g^{(1)}{}_{23}(\omega) = (C^2\eta^4)e^{i(q(x-\eta)+\omega)}(1-i/q\eta)/\eta^2 \tag{21}$$

The stress-energy tensor of the apparent matter associated with this polarization is identical to (19) and (20) and does not warrant separate discussion.

Orders of magnitude. The important factor in (19) and (20) is $\tilde{\varepsilon}(\eta)^2 n_{\mathrm{osc}}^2$. The energy density and pressure of apparent matter are (at this order) quadratic, not only in the amplitude $\tilde{\varepsilon}(\eta)$ of the perturbation, but also in its "frequency" n_{osc}. Thus, $\tilde{\varepsilon}(\eta) \ll 1$ does not necessarily translate into negligible energy density and pressure of apparent matter: the smallness of $\tilde{\varepsilon}(\eta)$ can be compensated by a sufficiently high frequency n_{osc}. For example, gravitational waves of relative amplitude $\tilde{\varepsilon}(\eta) \approx 10^{-5}$ and oscillation number $n_{\mathrm{osc}} \approx 10^5$ would generate an effective apparent large-scale stress-energy comparable to the energy density of the dust present in this model. Note that such a wave would have today a physical frequency of order 10^{-12} Hz and would thus elude direct observation [23].

3.3 Fluctuations in the density of matter

The first-order expressions for the metric and stress-energy tensors corresponding to a matter density fluctuation around a spatially flat, homogeneous and isotropic universe are well-known and given in [21]. These expressions can be used to compute the stress-energy of apparent matter from the formula (12).

We discuss here the simplest such perturbation; other types of density fluctuations are presented in the Appendix. The perturbation is of the form

$$\begin{aligned} {g^{(1)}}_{00} = 0 \qquad {g^{(1)}}_{11}(\omega) &= a(\eta)^2\eta^2\cos(qx+\omega) \\ {g^{(1)}}_{22}(\omega) = {g^{(1)}}_{33}(\omega) &= -a(\eta)^2\frac{10}{q^2}\cos(qx+\omega), \end{aligned} \tag{22}$$

corresponding to the following first-order stress-energy tensor perturbation

$$\mathcal{T}^{(1)}{}^0_0(\omega) = \rho\,\frac{\eta^2}{2}\cos(qx+\omega) \qquad \mathcal{T}^{(1)}{}^j_i = \mathcal{T}^{(1)}{}^j_0 = 0. \tag{23}$$

This stress-energy tensor describes a co-moving, sheer-free spatial density fluctuation. Note that the relative amplitude of the perturbation increases with time, which traces the aggregating effect of gravitation.

The quantity $\tilde{\varepsilon}(\eta) = \varepsilon\eta^2$ measures the effective relative magnitude of the perturbation $\varepsilon g^{(1)}$ with respect to g^{ref}. The quantity $n_{\mathrm{osc}} = q\eta$ represents the number of oscillations (periods) in that part of the universe accessible to an observer situated at time η; typically $n_{\mathrm{osc}} \gg 1$.

As above, the averaging is over $\omega \in [0; 2\pi]$, and the stress-energy tensor of apparent matter can be obtained from (12) by direct computation. This gives

$$\mathcal{T}^{\mathrm{app}}{}^0_0 = -\tilde{\varepsilon}(\eta)^2\frac{1-75/n_{\mathrm{osc}}^2}{16\pi}\rho(\eta) \tag{24}$$

$$\mathcal{T}^{\mathrm{app}}{}^1_1 = \tilde{\varepsilon}(\eta)^2\frac{25}{16\pi n_{\mathrm{osc}}^2}\rho(\eta) \tag{25}$$

$$\mathcal{T}^{\mathrm{app}}{}^2_2 = \mathcal{T}^{\mathrm{app}}{}^3_3 = \tilde{\varepsilon}(\eta)^2\frac{7+50/n_{\mathrm{osc}}^2}{32\pi}\rho(\eta) \tag{26}$$

and all the other terms are 0 at this order in ε.

The apparent matter associated with these fluctuations is thus characterized, at this order, by a negative energy density and a negative pressure. Loosely speaking, the negative energy could be interpreted in a semi-Newtonian setting as the gravitational energy of the fluctuations and the negative pressure represents the collapsing effects of gravitation.

There is an important difference w.r.t. the gravitational wave case above, namely that the effect simply scales like the square of the effective amplitude of the perturbation, with no n_{osc}^2 factor (compare (19) and (20)). Thus, the net large-scale effect of high-frequency gravitational waves is much more important than the net large-scale effect of matter density fluctuations of comparable wavelength, at least at this order in ε.

4 Conclusion

We have investigated how small-scale fluctuations influence the homogeneous and isotropic large-scale expansion of cosmological models. We have restricted the discussion to dust models and studied fluctuations in matter density as well as gravitational waves. Our perturbative results indicate that the so-called back reaction effect is dominated by gravitational waves, rather than matter density fluctuations. The relative importance of the effective large-scale stress-energy generated by gravitational waves scales as the squared product of their amplitude by their frequency. Thus, even small amplitude waves can generate an important effect provided their frequencies are high enough. For example, it is found that waves of current amplitude $\sim 10^{-5}$ and current physical frequency 10^{-12} Hz, which are not detectable with today's technology, would generate a large-scale stress-energy comparable to the dust energy.

The equation of state of the large-scale stress-energy generated by an isotropic background of gravitational waves is simply the equation of state of radiation with postive energy density and pressure. On the other hand, the matter density fluctuations we studied lead to negative energy density and pressure.

The results presented here prove that small-scale fluctuations can influence drastically the large-scale expansion of the universe and that back reaction cannot be *a priori* neglected in cosmology. One can then wonder if at least part of the cosmological dark energy cannot be interpreted as a large-scale signature of such small-scale fluctuations. The material presented in this article is not yet sufficient to reach a definitive conclusion in this matter. Let us nevertheless remark that the extremely simple cosmological models considered in this manuscript are already rich enough to generate apparent matters with very different equations of state, and that equations of state strongly ressembling that of the cosmological dark energy has been found by averaging a Schwarzschild black hole [18]. This work thus needs to be extended in several directions before a clear-cut conclusion can be reached. First, computations should be carried out on more general models than flat dust cosmologies. Second, the non-perturbative regime should be addressed, for example by numerical simulations. Third, different types of fluctuations should be combined and allowed to interfere with each other.

Acknowledgments. Thanks to Denis Serre for help with the curved-space gravitational wave equation.

Appendix: More on dust density fluctuations

First-order perturbations of a Friedmann–Lemaître–Robertson–Walker metric are described in [21] and are of various types. One of them is the gravitational wave considered in Section 3.2. For dust cosmologies, the next one reads:

$$\begin{aligned} {g^{(1)}}_{00} = 0 \qquad {g^{(1)}}_{11} &= -a(\eta)^2\beta(\eta)\cos(qx+\omega) \\ {g^{(1)}}_{22} = {g^{(1)}}_{33} &= -a(\eta)^2\gamma(\eta)\cos(qx+\omega)/q^2; \end{aligned} \tag{27}$$

associated with first-order fluctuations of matter

$$\mathcal{T}^{(1)}{}^{0}_{0} = \frac{\cos(qx+\omega)}{8\pi a^3}\left(a\gamma + 2\dot{a}\dot{\gamma}/q^2 + \dot{a}\dot{\beta}\right) \qquad \mathcal{T}^{(1)}{}^{1}_{0} = \frac{\dot{\gamma}\sin(qx+\omega)}{8\pi q a^2}, \tag{28}$$

the other components being 0. Here according to [21], the functions $\beta(\eta)$ and $\gamma(\eta)$ must satisfy

$$\ddot{\gamma} + 2\frac{\dot{a}}{a}\dot{\gamma} = 0 \qquad \ddot{\beta} + 2\frac{\dot{a}}{a}\dot{\beta} + \gamma = 0 \tag{29}$$

(The case given in the text is the simplest solution $\beta = -\eta^2, \gamma = 10$.)

Our expression (12) for apparent matter yields

$$\mathcal{T}^{\mathrm{app}}{}^{0}_{0} = \frac{\varepsilon^2}{64\pi a^2}\left(3\gamma^2/q^2 + 2\beta\gamma - \dot{\gamma}^2/q^4 - 2\dot{\beta}\dot{\gamma}/q^2 + 4\frac{\dot{a}}{a}\beta\dot{\beta} + 8\frac{\dot{a}}{a}\gamma\dot{\gamma}/q^4\right) \tag{30}$$

$$\mathcal{T}^{\mathrm{app}}{}^{1}_{1} = \frac{\varepsilon^2}{64\pi a^2}\left(\gamma^2/q^2 + \dot{\gamma}^2/q^4\right) \tag{31}$$

$$\mathcal{T}^{\mathrm{app}}{}^{2}_{2} = \mathcal{T}^{\mathrm{app}}{}^{3}_{3} = \frac{\varepsilon^2}{64\pi a^2}\left(\gamma^2/q^2 - \beta\gamma + \dot{\beta}^2 + \dot{\gamma}^2/q^4 - \dot{\beta}\dot{\gamma}/q^2\right) \tag{32}$$

and the other components are 0 or $O(\varepsilon^3)$.

In the regime we are interested in, $q \gg 1$, this reduces to

$$\mathcal{T}^{\mathrm{app}}{}^{0}_{0} = \frac{\varepsilon^2}{32\pi a^2}\left(\beta\gamma + 2\frac{\dot{a}}{a}\beta\dot{\beta}\right) \tag{33}$$

$$\mathcal{T}^{\mathrm{app}}{}^{1}_{1} = 0 \tag{34}$$

$$\mathcal{T}^{\mathrm{app}}{}^{2}_{2} = \mathcal{T}^{\mathrm{app}}{}^{3}_{3} = \frac{\varepsilon^2}{64\pi a^2}\left(\dot{\beta}^2 - \beta\gamma\right) \tag{35}$$

Since β and γ satisfy the second-order differential system (29), we can prescribe β, $\dot{\beta}$, γ and $\dot{\gamma}$ arbitrarily at one point in time. In particular, this leads to arbitrary signs for the energy and pressure of apparent matter. However, for large η, the system (29) implies that γ will tend to a constant and β will grow in time like η^2: this is the most interesting case, discussed in Section 3.3.

The last type of perturbation mentioned in [21] corresponds to a pure sheer perturbation; it takes the form

$$g^{(1)}{}_{12} = g^{(1)}{}_{21} = -a(\eta)^2\beta(\eta)\cos(qx+\omega)/q \tag{36}$$

corresponding to first-order stress-energy perturbation

$$\mathcal{T}^{(1)}{}^{2}_{0} = -\mathcal{T}^{(1)}{}^{0}_{2} = \frac{\dot{\beta}\sin(qx)}{2a^2} \tag{37}$$

where β satisfies $\ddot{\beta} + 2\frac{\dot{a}}{a}\dot{\beta} = 0$ i.e. $\beta = C_1/\eta^3$ in our case (the integration constant is a gauge choice).

For apparent matter this yields

$$\mathcal{T}^{\mathrm{app}\,0}_{\;\;0} = \frac{\varepsilon^2}{64\pi a^2 q^2}\left(\dot\beta^2 + 8\frac{\dot a}{a}\beta\dot\beta\right) \tag{38}$$

$$\mathcal{T}^{\mathrm{app}\,1}_{\;\;1} = \mathcal{T}^{\mathrm{app}\,2}_{\;\;2} = \frac{\varepsilon^2\dot\beta^2}{64\pi a^2 q^2} \qquad \mathcal{T}^{\mathrm{app}\,3}_{\;\;3} = \frac{3\varepsilon^2\dot\beta^2}{64\pi a^2 q^2} \tag{39}$$

So, not only do sheer perturbations decrease with time like $1/\eta^3$, but the apparent matter effect is small at high frequencies.

References

[1] Z. Burda, J. Jurkiewicz, and M.A. Nowak. Is Econophysics a solid Science? *Acta Phys. Pol. B*, 34 (1):87, 2003.

[2] J. Southern et al. Multi-scale computational modelling in biology and physiology. *Progr. Biophys. Mol. Bio.*, 96 (1–3):60–89, 2008.

[3] U. Frisch. *Turbulence.* Cambridge University Press, Cambridge, 1995.

[4] J. Cardy. *Scaling and renormalization in statistical physics.* Cambridge University Press, Cambridge, 1996.

[5] R.M. Wald. *General Relativity.* The University of Chicago Press, Chicago, 1984.

[6] S. Weinberg. *Cosmology.* Oxford University Press, 2008.

[7] F. Debbasch. What is a mean gravitational field? *Eur. Phys. J. B*, 37(2):257–270, 2004.

[8] F. Debbasch. Mean field theory and geodesics in general relativity. *Eur. Phys. J. B*, 43(1):143–154, 2005.

[9] T. Buchert. A cosmic equation of state for the inhomogeneous universe: can a global far-from-equilibrium state explain dark energy? *Class. Quantum Grav.*, 22(19):L113–L119, 2005.

[10] H. Alnes, M. Amarzguioui and Ø. Grøn. An inhomogeneous alternative to dark energy? *Phys. Rev. D*, 73:083519, 2006.

[11] T. Buchert. On globally static and stationary cosmologies with or without a cosmological constant and the dark energy problem. *Class. Quantum Grav.*, 23:817–844, 2006.

[12] E.W. Kolb, S. Matarrese, and A. Riotto. On cosmic acceleration without dark energy. *New J. Phys.*, 8:322, 2006.

[13] A. Paranjape and T.P. Singh. The possibility of cosmic acceleration via spatial averaging in Lemaître–Tolman–Bondi models. *Class. Quant. Grav.*, 23:6955–6969, 2006.

[14] S.Räsänen. Constraints on backreaction in dust universes. *Class. Quantum Grav.*, 23(6):1823–1835, 2006.

[15] T. Buchert. Dark energy from structure: a status report. *Gen. Rel. Grav.*, 40:467–527, 2008.

[16] C. Chevalier, F. Debbasch, and Y. Ollivier. Large scale nonlinear effects of fluctuations in relativistic gravitation. *Nonlinear Analysis Series A: Theory, Methods & Applications, In Press.*

[17] J. Larena, J.M. Alimi, T. Buchert, M. Kuntz, and P.S. Corasaniti. Testing backreaction effects with observations. *arXiv:0808.1161*, 2008.

[18] F. Debbasch and Y. Ollivier. Observing a Schwarzschild black-hole with finite precision. *Astron. Astrophys.*, 433(2):397–404, 2005.

[19] C. Chevalier, M.D. Bustamante, and F. Debbasch. Thermal statistical ensembles of black holes. *Physica A*, 376:293–307, 2007.

[20] C. Chevalier and F. Debbasch. Thermal statistical ensembles of classical extreme black holes. *Physica A*, 388:628–638, 2009.

[21] L.D. Landau and E.M. Lifshitz. *The Classical Theory of Fields.* Pergamon Press, Oxford, 4th edition, 1975.

[22] R.A. Isaacson. Gravitational radiation in the limit of high frequency. II. Nonlinear terms and the effective stress tensor. *Phys. Rev.*, 166:1272, 1968.

[23] W. Zhao and Y. Zhang. Relic gravitational waves and their detection. *Phys. Rev. D*, 74:043503, 2006.

Observing a Schwarzschild black hole with finite precision

Cet article écrit en 2004 est le premier de ma collaboration avec Fabrice Debbasch (Paris 6), qui s'est poursuivie dans le cadre du projet ANR que j'ai encadré en 2005–2008. Il est paru au volume 433, n°2 (2005) de la revue Astronomy and Astrophysics*, pp. 397–404.*

Observing a Schwarzschild black hole with finite precision

Fabrice Debbasch & Yann Ollivier

Abstract

We investigate how the space-time of a vacuum Schwarzschild black hole would appear if observed with a finite precision in the measurements of the spatial Kerr-Schild coordinates. For this we use the general procedure for evaluating mean gravitational fields recently presented in [Deb04b]. It is found that the black hole would then appear as surrounded by an apparent matter characterized by a negative energy density and two different pressures, a negative and a positive one. The total combined effect of the apparent matter leads to a space-time of negative scalar curvature, like de Sitter space-time. However, the 'magnitude' of the trace-free Ricci tensor does not vanish for this space-time, whereas it does for de Sitter space-time. Possible cosmological implications, concerning the evaluation of the mean density of the Universe and the cosmological constant, are also discussed.

Notation

In this article, space-time indices running from 0 to 3 will be indicated by Greek letters. The metric signature will be $(+,-,-,-)$. We also have chosen, as a rule, *not* to use the so-called intrinsic notation in differential geometry, but to use the notation standard in physics, which denotes each tensor by its components.

1 Introduction

Every observation is necessarily finite *i.e.* it deals with a finite number of quantities, observed or measured with a finite precision. This explains why mean field theories play such an important role in physics. It will therefore come as no surprise that developing a mean field approach to relativistic gravitation has been the subject of active research for more than a decade [Fut91, Fut93, Kas92, Fut96, Zal97, Buc00, Buc01]. This conceptually and practically crucial problem has been recently solved in a rather general way [Deb04b, Deb04a]. It has been shown that, given a statistical ensemble Σ of space-times sharing a common topology, it makes both mathematical and physical sense to define the mean (or apparent, or coarse-grained) space-time associated with this ensemble as a space-time of the same topology, but where the gravitational field is represented by a metric which is simply the average of the metrics corresponding to the various space-times members of Σ.

This apparently very innocuous result has however several exotic consequences. One of them is that the separation between the gravitational field and the matter degrees of freedom actually depends on the precision of the observations [Deb04b]. Let us consider the following particular situation. Suppose a region $\mathcal{D}$ of space-time is observed with a certain finite precision and that the observations indicate that no matter is present in $\mathcal{D}$, but only a non-vanishing gravitational field. Then, generically, other observations carried out with a different (greater or lesser) precision will indicate that $\mathcal{D}$ contains both matter and a non-vanishing gravitational field. The aim of this article is to investigate this 'purely relativistic' effect on a perhaps academic but *de facto* simple and hopefully illuminating example, where most calculations can be made completely explicit. More precisely, we consider the Schwarzschild black hole, which is one of the simplest vacuum solutions to Einstein's equation and we study how a finite precision in coordinate measurements can make it look like a non-vacuum solution to the general relativistic field equations.

The material is organized as follows. Section 2 reviews some basic results about ensembles of space-times and about the properties of the mean or coarse-grained gravitational field with which they are associated. Section 3 introduces the particular statistical ensemble which will be considered in this article; it is notably explained why averaging over this statistical ensemble can be interpreted as observing a Schwarzschild black hole with a finite precision. In Section 4, we calculate the mean metric associated with this statistical ensemble and, in Section 5, the stress-energy tensor of the apparent matter which seems to surround the black hole is explicitly evaluated as a function of the coarse-graining; the calculation is a perturbative one and is valid for points whose radial (Schwarzschild) coordinates are much larger than the coarse-graining itself. At lowest order, it is found that the apparent matter can be characterized by an energy density and two different pressures; the energy density and one of the pressures is negative, while the other pressure is positive. All three quantities decrease towards zero as the radial coordinates tends to infinity. We also show that the total effect of this apparent matter is to induce a negative scalar curvature in space-time. Thus, by coarse-graining, the vacuum surrounding the Schwarzschild black hole acquires a stress-energy tensor which generates a space-time of negative curvature. This obviously brings to mind de Sitter space-time, the negative curvature of which is generated by a non-vanishing positive cosmological constant. Section 6 provides an in-depth discussion of the results presented in this article, including possible cosmological implications. In particular, the similarities and differences between the apparent vacuum stress-energy due to the coarse graining and the stress-energy corresponding to a cosmological constant are analyzed. As a conclusion, we provide a summary of the new material and we also mention and discuss briefly some of the many possible extensions of this work, including several more realistic situations of direct astrophysical and/or cosmological relevance.

2 Mean gravitational fields

2.1 Ensembles of space-times

Let us consider a statistical ensemble Σ of space-times $\mathcal{M}(\omega)$, $\omega \in \Omega$. Ω is an arbitrary probability space [GS94]; each member of the ensemble Σ is a differentiable manifold endowed with a metric $g(\omega)$, the Levi-Civita connection $\Gamma(\omega)$ associated with $g(\omega)$ [Nak90] and a stress-energy tensor $T(\omega)$.

We will restrict the discussion by supposing that all space-times in our statistical ensemble share the same topology and are distinguished only by their respective gravitational fields. More precisely, we suppose that there is a single manifold M underlying all of our space-times $\mathcal{M}(\omega)$ (such an M represents the set of points of space-time), so that $\mathcal{M}(\omega)$ is M equipped with an ω-dependent metric field $g(\omega)$. One can thus choose an atlas common to all space-times, so that for any chart (i.e. any local coordinate system (x)), $\mathcal{M}(\omega)$ is represented by an ω-dependent metric field $g_{\mu\nu}(x,\omega)$.

Each space-time $\mathcal{M}(\omega)$ verifies the Einstein equation [Wal84]. One thus has:

$$\mathcal{E}_{\mu\nu}(g(\omega)) \equiv R_{\mu\nu}(\omega) - \frac{1}{2} R(\omega) g_{\mu\nu}(\omega) = \chi g_{\mu\alpha}(\omega) g_{\nu\beta}(\omega)\, T^{\alpha\beta}(\omega), \tag{1}$$

where the $R_{\mu\nu}$'s are the coordinate-basis components of the Ricci tensor, R is the trace of this tensor and χ is the gravitational constant. The combination on the left-hand side of (1) is usually called the Einstein tensor, hence the notation. Unless otherwise specified, the units used in the rest of this article are so chosen that $\chi = 8\pi$ [Wal84].

2.2 Definition of a mean space-time

It has been shown in [Deb04b] that the statistical ensemble Σ of space-times can be used to define a single, mean Einstein space-time $\bar{\mathcal{M}}$ and that, by construction, the atlas common to all members of Σ can be used as an atlas for $\bar{\mathcal{M}}$. $\bar{\mathcal{M}}$ is endowed with a metric $\bar{g}$ which is the average of the metrics $g(\omega)$ over ω; one thus has, for all x:

$$\bar{g}(x) = \langle g(x,\omega) \rangle, \tag{2}$$

where the brackets on the right-hand side indicate an average over the statistical ensemble Σ.

The connection of the mean space-time $\bar{\mathcal{M}}$ is simply the Levi-Civita connection associated with the metric $\bar{g}$ and will be conveniently called the mean connection. Since the relations linking the components $g_{\mu\nu}$ of an arbitrary metric g to the Christoffel symbols $\Gamma^{\alpha}_{\mu\nu}$ of its Levi-Civita connection are non-linear, the Christoffel symbols of the mean connection are *not* identical to the averages of the Christoffel symbols associated with the various space-times $\mathcal{M}(\omega)$. Note however that the so-called 'covariant' connection coefficients $\Gamma_{\mu,\alpha\beta}(\omega) \equiv g_{\mu\nu}(\omega)\Gamma^{\nu}_{\alpha\beta}(\omega)$ depend linearly on the metric components $g_{\mu\nu}(\omega)$, so that $\bar{\Gamma}_{\mu,\alpha\beta} = \langle \Gamma_{\mu,\alpha\beta}(\omega) \rangle$. This point is thoroughly elaborated upon in [Deb04b], where a complete discussion of the mathematical and physical motivations for definition (2) can also be found.

Because the Einstein tensor depends non-linearly on the metric, the Einstein tensor $\bar{\mathcal{E}} = \mathcal{E}(\bar{g})$ associated with the mean metric *does not* generally coincide with the average of the Einstein tensors $\mathcal{E}(g(\omega))$. The tensor $\bar{\mathcal{E}}$ is nevertheless the Einstein tensor of the mean space-time. It therefore defines, via the Einstein equation, a stress-energy tensor $\bar{T}$ for the mean space-time:

$$\mathcal{E}_{\mu\nu}(\bar{g}) = \chi \bar{g}_{\mu\alpha} \bar{g}_{\nu\beta} \bar{T}^{\alpha\beta}. \tag{3}$$

Since $\mathcal{E}_{\mu\nu}(\bar{g}) \neq \langle \mathcal{E}_{\mu\nu}(g(\omega)) \rangle$, the mean stress-energy tensor $\bar{T}^{\alpha\beta}$ is generally different from the average $\langle T^{\alpha\beta}(\omega) \rangle$ of the stress-energy tensors of the space-times in the statistical distribution. It is therefore convenient to introduce the generally non-vanishing tensor field ΔT, defined on $\bar{\mathcal{M}}$ by :

$$\Delta T^{\alpha\beta} = \bar{T}^{\alpha\beta} - \left\langle T^{\alpha\beta}(\omega) \right\rangle. \tag{4}$$

This difference ΔT can be interpreted as the stress-energy tensor of an "apparent matter" which contributes, along with the average $\langle T(\omega) \rangle$ of the stress-energy associated with the 'real' matter present in the various original space-times $\mathcal{M}(\omega)$, to creating the mean gravitationnal field $\bar{g}$:

$$\mathcal{E}_{\mu\nu}(\bar{g}) = \chi \bar{g}_{\mu\alpha} \bar{g}_{\nu\beta} \left(\left\langle T^{\alpha\beta}(\omega) \right\rangle + \Delta T^{\alpha\beta} \right). \tag{5}$$

In particular, the vanishing of $T(\omega)$ for all ω does not necessarily imply the vanishing of $\bar{T}$. The mean stress-energy tensor $\bar{T}$ can therefore be non-vanishing in regions where the unaveraged stress-energy tensor actually vanishes. A general dicussion of this and other perhaps unexpected consequences of definition (2) can be found in [Deb04b, Deb04a]. The particular cases when the matter is made of an electromagnetic field and/or of a possibly charged perfect fluid is also addressed in depth by [Deb04b].

The goal of this article is to present a simple case when $T(\omega)$ vanishes for all ω and ΔT is nevertheless non-zero.

3 An ensemble of space-times representing a Schwarzschild black hole observed with a finite precision

The so-called Kerr-Schild form [Cha92, KSMH80] of the (vacuum) Schwarzschild metric is :

$$ds^2 = dt^2 - d\mathbf{r}^2 - \frac{2M}{r} \left(dt + \frac{\mathbf{r}}{r} \cdot d\mathbf{r} \right)^2. \tag{6}$$

The parameter M represents the mass of the black hole and $\mathbf{r}$ stands for the set of three 'spatial' coordinates x, y, z. We have also retained the standard and natural notations:

$$d\mathbf{r}^2 = dx^2 + dy^2 + dz^2 \tag{7}$$

and

$$\mathbf{r} \cdot d\mathbf{r} = x dx + y dy + z dz. \tag{8}$$

The Kerr-Schild coordinates (as opposed to the perhaps more standard Schwarzschild coordinates [Wal84]) are particularly natural and convenient because they form a single-chart atlas of the whole Schwarzschild space-time [Cha92], the only singularities of this space-time being the points on the 'line' $x = y = z = 0$, where the components of the metric tensor (6) are themselves singular.

Let us now introduce an at this stage arbitrary $\boldsymbol{\omega}$ in $\mathbb{R}^3$ and consider the $\boldsymbol{\omega}$-dependent metric

$$ds^2_{\boldsymbol{\omega}} = dt^2 - d\mathbf{r}^2 - \frac{2M}{\rho}\left(dt + \frac{\mathbf{r}-\boldsymbol{\omega}}{\rho}\cdot d\mathbf{r}\right)^2, \tag{9}$$

where

$$\rho^2 = (\mathbf{r}-\boldsymbol{\omega})^2 = r^2 + \omega^2 - 2\mathbf{r}\cdot\boldsymbol{\omega}. \tag{10}$$

Note that the 'original' Schwarzschild space-time associated with (6) is actually $\mathcal{M}(0)$; also observe that, for any $(t, \mathbf{r})$ and $\boldsymbol{\omega}$:

$$g(t, \mathbf{r}, \boldsymbol{\omega}) = g(t, \mathbf{r}-\boldsymbol{\omega}, 0), \tag{11}$$

so that $g(t, \mathbf{r}, \boldsymbol{\omega})$ represents an ordinary black hole centered around point $\boldsymbol{\omega}$.

Let $\Omega = \{\boldsymbol{\omega} \in \mathbb{R}^3; \omega^2 \leqslant a^2\}$ where a is a fixed, positive real constant; Ω is the usual 3-ball of radius a in Euclidean space $\mathbb{R}^3$. We will use as volume measure on Ω the usual (Lebesgue) measure $d^3\boldsymbol{\omega}$ and, with this measure, the total volume of Ω is simply $V_a = 4\pi a^3/3$. The measure $d^3\boldsymbol{\omega}$ thus defines a probability measure on Ω by :

$$p(\boldsymbol{\omega})d^3\boldsymbol{\omega} = \frac{1}{V_a}d^3\boldsymbol{\omega}. \tag{12}$$

We now define a statistical ensemble Σ of space-times by $\Sigma = \{\mathcal{M}(\boldsymbol{\omega}); \boldsymbol{\omega} \in \Omega\}$ and use on Σ the probability measure (12).

The remainder of this article is devoted to investigating some properties of the average space-time $\bar{\mathcal{M}}$ which can be constructed out of this ensemble by the procedure outlined in the previous section. Before embarking on any calculation, let us give a physical motivation for considering the ensemble Σ.

At any point $(t, \mathbf{r})$ in space-time, the value $\bar{g}(t, \mathbf{r})$ taken by the metric $\bar{g}$ of the average space-time $\bar{\mathcal{M}}$ is simply the average of $g(t, \mathbf{r}, \boldsymbol{\omega})$ over $\boldsymbol{\omega}$. One thus has, by equation (11) :

$$\bar{g}(t, \mathbf{r}) = \langle g(t, \mathbf{r}-\boldsymbol{\omega}, 0)\rangle. \tag{13}$$

This shows that, at any given point $(t, \mathbf{r})$ in space-time, the metric $\bar{g}$ is simply the average of the original metric (6) over the 3-ball of radius a centered at $(t, \mathbf{r})$.

The metric $\bar{g}$ can therefore be interpreted as the original metric $g(0)$ observed, in the chosen coordinates, with the finite 'spatial' resolution a. It thus represents a Schwarzschild black hole observed with a finite precision. Indeed, suppose that, by some observational procedure, we can have experimental access to the metric tensor field g but suppose also that the determination of each 'spatial' Kerr-Schild coordinate is subject to an error of order a. Then, instead of measuring, say $g(t, \mathbf{r})$, we actually measure $g(t, \mathbf{r} - \boldsymbol{\omega})$ for some randomly chosen $\boldsymbol{\omega}$ of norm at most a (in the sense of equation 7). The 'observed' or 'measured' metric will then precisely be $\bar{g}(t, \mathbf{r}) = \langle g(t, \mathbf{r} - \boldsymbol{\omega}, 0) \rangle$.

As explained in the previous section, the average metric $\bar{g}$ defines by Einstein's equation an *a priori* non-vanishing stress-energy tensor $\bar{T}$. In other words, although each metric $g(t, \mathbf{r}, \boldsymbol{\omega})$ in the ensemble Σ is a solution of Einstein's equation in vacuum, the average metric $\bar{g}$ is *not*. If measurements are made with a finite 'spatial' resolution a, the observed metric $\bar{g}$ can only be consistently understood as a solution of Einstein's equation if one takes into account an 'apparent' matter caracterized by the stress-energy tensor $\bar{T}$. We now want to investigate the properties of this matter in greater detail.

4 Determination of the mean metric

4.1 Kerr-Schild coordinates

We first begin by determining the average metric $\bar{g}$, fully defined by $\bar{g}(t, \mathbf{r}) = \langle g(t, \mathbf{r} - \boldsymbol{\omega}, 0) \rangle$. For obvious physical reasons, one is only interested in evaluating the mean metric $\bar{g}$ at points $(t, \mathbf{r})$ for which $r \gg a$. This we will now do, pushing all expansions at order two in a/r.

Equation (9) can be rewritten as:

$$ds^2 = dt^2 - d\mathbf{r}^2 - \frac{2M}{\rho}\left(dt^2 + \frac{2dt}{\rho} d\mathbf{r} \cdot (\mathbf{r} - \boldsymbol{\omega}) + \frac{1}{\rho^2}\left(d\mathbf{r} \cdot (\mathbf{r} - \boldsymbol{\omega})\right)^2\right) \tag{14}$$

where as above, $\rho^2 = (\mathbf{r} - \boldsymbol{\omega})^2$.

To proceed, one needs to expand the various powers of $1/\rho$ which enter (14) into powers of $\mathbf{r}$ and $\boldsymbol{\omega}/\mathbf{r}$.

The powers of $1/\rho$. Here we begin to use the assumption that $r \gg a$. All subsequents expansions are at order 2 in ω/r.

Expanding $1/\rho = 1/\sqrt{r^2 + \omega^2 - 2\mathbf{r} \cdot \boldsymbol{\omega}}$ at order 2 in ω/r we get

$$\frac{1}{\rho} = \frac{1}{r}\left(1 + \frac{\mathbf{r} \cdot \boldsymbol{\omega}}{r^2} - \frac{1}{2}\frac{\omega^2}{r^2} + \frac{3}{2}\frac{(\mathbf{r} \cdot \boldsymbol{\omega})^2}{r^4}\right) \tag{15}$$

$$\frac{1}{\rho^2} = \frac{1}{r^2}\left(1 + 2\frac{\mathbf{r} \cdot \boldsymbol{\omega}}{r^2} - \frac{\omega^2}{r^2} + 4\frac{(\mathbf{r} \cdot \boldsymbol{\omega})^2}{r^4}\right) \tag{16}$$

$$\frac{1}{\rho^3} = \frac{1}{r^3}\left(1 + 3\frac{\mathbf{r} \cdot \boldsymbol{\omega}}{r^2} - \frac{3}{2}\frac{\omega^2}{r^2} + \frac{15}{2}\frac{(\mathbf{r} \cdot \boldsymbol{\omega})^2}{r^4}\right) \tag{17}$$

Intermediate forms of the mean metric. We now plug these expansions into equation (14) and average for $\boldsymbol{\omega}$ in the ball Ω of radius a. By symmetry, it is clear that the average of $\boldsymbol{\omega}$ is 0, as well as the average of all terms containing an odd power of $\boldsymbol{\omega}$. We get

$$\begin{aligned}\left\langle ds^2\right\rangle &= dt^2 - d\mathbf{r}^2 - \frac{2M}{r}dt^2\left(1-\frac{1}{2r^2}\langle\omega^2\rangle+\frac{3}{2r^4}\left\langle(\mathbf{r}\cdot\boldsymbol{\omega})^2\right\rangle\right)\\
&\quad-\frac{4M}{r^2}dt\,d\mathbf{r}\cdot\left(\mathbf{r}\left(1-\frac{1}{r^2}\langle\omega^2\rangle+\frac{4}{r^4}\left\langle(\mathbf{r}\cdot\boldsymbol{\omega})^2\right\rangle\right)-\frac{2}{r^2}\langle\boldsymbol{\omega}(\mathbf{r}\cdot\boldsymbol{\omega})\rangle\right)\\
&\quad-\frac{2M}{r^3}\left(\begin{array}{l}\left\langle(d\mathbf{r}\cdot(\mathbf{r}-\boldsymbol{\omega}))^2\right\rangle+\frac{3}{r^2}\left\langle(\mathbf{r}\cdot\boldsymbol{\omega})\,(d\mathbf{r}\cdot(\mathbf{r}-\boldsymbol{\omega}))^2\right\rangle\\ \qquad-\frac{3}{2r^2}\left\langle\omega^2\,(d\mathbf{r}\cdot(\mathbf{r}-\boldsymbol{\omega}))^2\right\rangle+\frac{15}{2r^4}\left\langle(d\mathbf{r}\cdot(\mathbf{r}-\boldsymbol{\omega}))^2\,(\mathbf{r}\cdot\boldsymbol{\omega})^2\right\rangle\end{array}\right)\end{aligned} \tag{18}$$

We thus need to compute the averages of several functions of $\mathbf{r}$ and $\boldsymbol{\omega}$. Symmetry arguments make the task easier. Remember that the average is taken on the Euclidean 3-ball of radius a. Since a is supposed to be much smaller than r, we only keep order-2 terms in a.

$$\left\langle\omega^2\right\rangle = \frac{3a^2}{5} \quad;\quad \left\langle(\mathbf{r}\cdot\boldsymbol{\omega})^2\right\rangle = \frac{a^2r^2}{5} \quad;\quad \langle\boldsymbol{\omega}(\mathbf{r}\cdot\boldsymbol{\omega})\rangle = \frac{a^2}{5}\mathbf{r} \tag{19}$$

$$\left\langle(d\mathbf{r}\cdot(\mathbf{r}-\boldsymbol{\omega}))^2\right\rangle = r^2dr^2+\frac{a^2}{5}\,d\mathbf{r}^2 \tag{20}$$

$$\left\langle(\mathbf{r}\cdot\boldsymbol{\omega})\,(d\mathbf{r}\cdot(\mathbf{r}-\boldsymbol{\omega}))^2\right\rangle = -\frac{2a^2r^2}{5}\,dr^2 \tag{21}$$

$$\left\langle\omega^2\,(d\mathbf{r}\cdot(\mathbf{r}-\boldsymbol{\omega}))^2\right\rangle = \frac{3a^2r^2}{5}\,dr^2 + \text{higher-order terms} \tag{22}$$

$$\left\langle(\mathbf{r}\cdot\boldsymbol{\omega})^2\,(d\mathbf{r}\cdot(\mathbf{r}-\boldsymbol{\omega}))^2\right\rangle = \frac{a^2r^4}{5}\,dr^2 + \text{higher-order terms} \tag{23}$$

Plugging this into expression (18) for $\left\langle ds^2\right\rangle$ we get the somewhat simpler form

$$\begin{aligned}\left\langle ds^2\right\rangle &= dt^2 - d\mathbf{r}^2\left(1+\frac{2Ma^2}{5r^3}\right)-\frac{2M}{r}\,dt^2\\
&\quad-\frac{4M}{r}\,dt\,dr\left(1-\frac{a^2}{5r^2}\right)-\frac{2M}{r}\,dr^2\left(1-\frac{3a^2}{5r^2}\right),\end{aligned} \tag{24}$$

which is the expression of the mean metric for a Schwarzschild black hole observed, in the retained coordinate system, with finite 'spatial' resolution a. Of course $a = 0$ gives back the usual metric (6). The deformation is of second order in a due to the symmetry of the ensemble Σ.

Comparing (24) to (6), we see that the role played by $d\mathbf{r}^2$ in (6) is now played by $d\mathbf{r}^2\left(1+\frac{2Ma^2}{5r^3}\right)$. This suggests the introduction of the new 'spatial' coordinates $\mathbf{R}$, defined by:

$$\mathbf{R} = \mathbf{r}\left(1+\frac{Ma^2}{5r^3}\right); \tag{25}$$

observe that $\mathbf{R}$ is equivalent to $\mathbf{r}$ at infinity.

Expressing $\langle ds^2\rangle$ in terms of the coordinates $(t,\mathbf{R})$ yields:

$$\langle ds^2\rangle = dt^2 - d\mathbf{R}^2 - \frac{2M}{R}\,dt^2\left(1+\frac{Ma^2}{5R^3}\right) - \frac{4M}{R}\,dt\,dR\left(1+\left(\frac{3M}{R}-1\right)\frac{a^2}{5R^2}\right) - \frac{2M}{R}\,dR^2\left(1+\frac{Ma^2}{R^3}\right), \tag{26}$$

or, equivalently,

$$\langle ds^2\rangle = dt^2 - d\mathbf{R}^2 - \frac{2M}{R}\left(dt\left(1+\frac{Ma^2}{10R^3}\right) + dR\left(1+\frac{Ma^2}{2R^3}\right)\right)^2 + \frac{4Ma^2}{5R^3}\,dt\,dR. \tag{27}$$

Both above expressions are correct at order 2 in a/R. They represent the finite-resolution version of (6). By a slight extension of the common terminology, we will say that the coordinates $(t,\mathbf{R})$ are Kerr-Schild coordinates for the average space-time $\bar{\mathcal{M}}$. Formally speaking, the only singularities of the metric (26) are the points on the line $R=0$. Thus, the Kerr-Schild coordinates form a single-chart atlas of the space-time equipped with metric (26). Note however that (26) was derived from (6) under the assumption that $r \gg a$, which implies $R \gg a$ via (25). The singularity of (26) at $R=0$ is therefore not 'physical', *i.e.* it does not entail that the mean space-time $\bar{\mathcal{M}}$ is singular at $R=0$. Moreover, the very notion of a mean space-time probably makes no physical sense for values of R comparable or inferior to the coarse graining parameter a[1].

4.2 Schwarzschild coordinates

By suitably choosing a new time-variable $\tau(t,\mathbf{r})$, the metric (6) can be put into the well-known form [Wal84, Cha92]:

$$ds^2 = \left(1-\frac{2M}{r}\right)d\tau^2 - \frac{1}{1-2M/r}\,dr^2 - r^2 d\Gamma^2, \tag{28}$$

where $d\Gamma$ stands for the elementary solid angle associated with the three 'spatial' coordinates $\mathbf{r}$. The coordinates $(\tau,\mathbf{r})$ are called the Schwarzschild coordinates. As already mentioned, these coordinates do not constitute a single-chart atlas of the Schwarzschild space-time. They are however relevant for observers outside the black hole and, if only because no term in $drdt$ appears in (28), they also present undeniable computational advantages. It is therefore natural to seek a new coordinate system

[1] Just as it makes no physical sense, for example, to speak of the electric field created by an electrostatic dipole at distances comparable or inferior to the caracteristic spatial extension of the charge distribution modelled by the dipole.

which would play for the mean metric (26) the role the usual Schwarzschild coordinate system plays for vacuum black holes. By extension, these new coordinates will be called the Schwarzschild coordinates of the mean metric.

They can be obtained by keeping $\mathbf{R}$ as 'spatial' coordinates and by merely introducing a new time-coordinate T, defined in terms of t and $\mathbf{R}$ by a relation of the form :

$$dt = dT + \alpha(R)dR \tag{29}$$

In fact it is not even necessary to compute $\alpha(R)$ explicitly to obtain the expression of the mean metric in Schwarzschild coordinates: indeed, the transformation (29) does not change the determinant of the metric (since this transformation is of determinant 1), and it does not change the term in front of dT^2 either. A simple computation shows that the determinant of the (t, R)-part of the metric (26) is -1 at order 2 in a/R. So the final metric will be of determinant -1 and, therefore, the terms in front of dT^2 and dR^2 will be the inverse of each other. Since the g_{TT} component is known from (26), so is the g_{RR} component. Naturally, this simple reasoning can be double-checked through a straightforward but rather long direct computation of $\alpha(R)$. Indeed, the choice :

$$\alpha(R) = \frac{2M}{R}\frac{1}{1-2M/R}\left(1 - \frac{a^2}{5R^2}\frac{1}{1-2M/R}\left(\frac{4M^2}{R^2} - \frac{5M}{R} + 1\right)\right) \tag{30}$$

ensures the vanishing of the mixed metric component g_{TR} and the g_{RR} component of the metric in Schwarzschild coordinates can then be obtained by direct computation.

The final form of the mean metric in Schwarzschild coordinates is therefore (with $d\Gamma$ the usual Euclidean solid angle element):

$$\langle ds^2 \rangle = F(R)dT^2 - G(R)dR^2 - R^2 d\Gamma^2 \tag{31}$$

where

$$F(R) = 1 - \frac{2M}{R} - \frac{2a^2M^2}{5R^4} \tag{32}$$

and

$$G(R) = \frac{1}{F(R)} \tag{33}$$

or, when $R - 2M \gg a^2M^2/R^3$,

$$G(R) = \frac{1}{1-2M/R} + \frac{2a^2M^2}{5R^4}\left(\frac{1}{1-2M/R}\right)^2 \tag{34}$$

There are a few simple but important remarks to be made about this result.

1. Expression (30) shows that the Schwarzschild coordinates cannot be used in the whole space-time $\bar{\mathcal{M}}$ but are only valid in the domains $R > 2M$ and $R < 2M$. On the other hand, the Kerr-Schild coordinates do form a single-chart atlas of $\bar{\mathcal{M}}$.

2. Of course, when $a = 0$, the preceding expression reduces to the metric (28) of a vaccum, non-rotating black hole in Schwarzschild coordinates. The average space-time is also static [Wal84], as can be seen from the absence of $dTdR$ term in (31). This was predictable since the mean space-time results from averaging static space-times.

3. The deformation of (31) with respect to (28) is of second order in a/R, due to symmetry of our statistical ensemble.

4. As noted above, the determinant of the (T, R)-part of the metric is -1:

$$F(R) = 1/G(R) \tag{35}$$

at order 2 in a/R. This property of the mean metric is shared by the metrics of the unaveraged space-times in Schwarzschild coordinates.

5. The only assumption that was made is that the coarse-graining a is much smaller than R. In particular, we have *not* assumed that R is large compared to the Schwarzschild radius M of the black hole: if the uncertainty a is small compared to M, then our estimate is valid even for $R \approx 2M$ *i.e.*, very close to the horizon of the unaveraged Schwarzschild black hole.

6. The metric components in the coordinate basis associated with T, R, θ and φ exhibit singularities for two different values of R. The singularity at $R = 0$ also appears in the form (27) of the average metric (see discussion above). As in the Schwarzschild case, this is a real singularity of the geometry defined by the metric (31); but let us stress again that the very notion of a mean space-time probably makes no physical sense for values of R comparable or inferior to the coarse graining parameter a. The other singularity of (31) occurs for $F(R) = 0$, namely $5R^4 - 10MR^3 - 2a^2M^2 = 0$. Solving this equation at order 2 in a/M (which is the order at which the mean metric has been computed), one finds that this singularity occurs at $R = 2M\left(1 + a^2/40M^2\right)$. This is a mere coordinate singularity, the occurence of which parallels the presence of an apparent singularity at $R = 2M$ for the components of Schwarzschild metric (28) in the basis associated with the usual Schwarzschild coordinates. The fact that the singularity at $R = 2M\left(1 + a^2/40M^2\right)$ is only due to the choice of coordinates in (31) can be checked in two different ways. First, this singularity is absent from the metric components (27) in the coordinate basis associatedwith (t, R, θ, φ). Second, the curvature tensor associated with (31) is well behaved everywhere, except at $R = 0$. In particular, the scalar curvature $\mathcal{R}$ of the mean space-time will be calculated in the next Section and is given by (39). It scales as R^{-6} and is everywhere finite, except at $R = 0$. The apparent singularity is thus due to the singular behaviour of the coordinate change defined by (29) and (30) at $R = 2M\left(1 + a^2/40M^2\right)$.

5 Energy and pressure of apparent matter

Let us insist that the mean metric (31) is *not* a solution of Einstein's equation in vacuum. We now wish to evaluate the stress-energy tensor $\bar{T}$ corresponding to this metric. This stress-energy tensor is the one that would be inferred by an observer having access to the coarse-grained metric (31); it therefore constitutes *apparent matter* that would be 'detected' by any observer mapping the gravitational field with the finite spatial precision a (in Kerr-Schild coordinates). We will restrict our discussion by investigating the properties of this apparent matter in the region $R > 2M$ only (the stress-energy tensor $\bar{T}$ in the region $R < 2M$ can be deduced similarly).

The use of Schwarzschild coordinates in the region $R > 2M$ allows for a very easy computation of $\bar{T}$. Indeed, for metrics of the form $e^\nu dT^2 - e^\lambda dR^2 - R^2 d\Gamma^2$ the stress-energy tensor can be readily expressed in terms of λ and ν (see for example [LL75], equations (100,2), (100,4), (100,6), or [Wal84]). In the case at hand, the calculation further simplifies since, first, $\lambda = -\nu$ (since $F = 1/G$) and, second, all functions are independent of the time coordinate T. One thus immediately gets:

$$8\pi\,\bar{T}_0^0 = -\,\frac{6a^2M^2}{5R^6}\,; \qquad 8\pi\,\bar{T}_1^1 = -\,\frac{6a^2M^2}{5R^6}\,; \qquad 8\pi\,\bar{T}_2^2 = \frac{12a^2M^2}{5R^6}\,; \qquad 8\pi\,\bar{T}_3^3 = \frac{12a^2M^2}{5R^6}; \tag{36}$$

all other components of $\bar{T}$ vanish, so that the stress-energy tensor $\bar{T}$ is diagonal in Schwarzschild coordinates.

As is well-known [LL75, Wal84], the component $\bar{T}_0^0$ can be interpreted as an energy density ε; in the present case, ε represents the energy-density of the apparent matter and the pressure of this matter in direction i is similarly given by $-T_i^i$.
We thus have:

$$\varepsilon = -\,\frac{1}{8\pi}\,\frac{6a^2M^2}{5R^6}\,; \qquad p_1 = \frac{1}{8\pi}\,\frac{6a^2M^2}{5R^6}\,; \qquad p_2 = -\,\frac{1}{8\pi}\,\frac{12a^2M^2}{5R^6}\,; \qquad p_3 = -\,\frac{1}{8\pi}\,\frac{12a^2M^2}{5R^6} \tag{37}$$

In particular, the apparent energy density is *negative*, and the pressure tensor is anisotropic; the radial direction (pointing towards the center of the black hole) is associated with a positive pressure whereas the single pressure associated with both perpendicular directions is negative.

It is also interesting to evaluate the (scalar) curvature $\mathcal{R}$ of the mean space-time; Einstein's equation (1) leads directly to:

$$R_\mu^\mu - \frac{1}{2}\,\mathcal{R} g_\mu^\mu = -\mathcal{R} = \chi\,(\varepsilon - p_1 - p_2 - p_3) \tag{38}$$

with $\chi = 8\pi$ in the chosen units, so that:

$$\mathcal{R} = -\,\frac{12a^2M^2}{5R^6}. \tag{39}$$

The averaging procedure thus confers on the space-time an apparent, strictly negative scalar curvature. In other words, after coarse-graining, the vacuum of the original Schwarzschild black hole appears endowed with a negative scalar curvature. This striking conclusion cannot but bring to mind the recent observation [S$^+$03] of a positive, non-vanishing cosmological constant Λ, which also endows vacuum regions of space-time with a negative scalar curvature [HE73, KT90, Pee93] $\mathcal{R}_\Lambda = -4\Lambda \times 8\pi$. This point will be further discussed below.

6 Discussion

Physical interpretation, for black holes, of the retained averaging. First, we discuss again the physical significance of the statistical ensemble of space-times Σ introduced in Section 4. As argued in that section, averaging the metric over this statistical ensemble leads to a new, mean metric $\bar{g}$ which represents the gravitationnal field 'detected' by someone who observes a vacuum Schwarzschild black hole with a finite precision a in the measurements of the 'spatial' Kerr-Schild coordinates. The point we would like to stress here is that the mean metric $\bar{g}$ does not represent the gravitational field detected by someone who observes a Schwarzschild black hole with finite precision a in the measurements of other coordinates, *e.g.* the 'spatial' *Schwarzschild* coordinates. To obtain the mean metric $\bar{g}'$ in that latter case, one would have to start with a new ensemble of space-times Σ', constructed from (28) exactly as Σ is constructed from (6), and evaluate $\bar{g}'$ as an average over this new ensemble Σ'[2]. There is obviously no reason why the metric $\bar{g}'$ should be identical to (28). Indeed, expression (28) represents the metric $\bar{g}$ (not $\bar{g}$), but in a coordinate system different from the one used in (26).

As for $\bar{g}'$, it can also be expressed in various coordinate systems. In one of them, which one would be entitled to call the Schwarzschild coordinate system for $\bar{g}'$, this metric would take a form similar to (28), but its expression would involve two functions *a priori* different from the functions F and G introduced in (31). And, extending these coordinates beyond $R = 2M$, one could probably construct a system of Kerr-Schild-like coordinates for $\bar{g}'$ too, where the metric resembles (26); but the expression of $\bar{g}'$ in these coordinates would not coincide with (26).

One might then wonder why we chose to evaluate the mean metric corresponding to a finite precision in the measurements of the spatial Kerr-Schild coordinates, and not the Schwarzschild ones. The reason is twofold. First, from a geometrical point of view, the use of Kerr-Schild coordinates is more natural because, as repeatedly stated, these coordinates constitute a single-chart atlas of the space-time manifold describing a Schwarzschild black hole, wheras the Schwarzschild coordinates are only valid outside (or inside) the horizon. As a consequence, practically any discussion of physics around a black hole is made easier by the use of Kerr-Schild coordinates. This relative simplicity will be used in subsequent publications, where the properties of the metric $\bar{g}$ ((26)) will be further investigated. The other mean metric $\bar{g}'$ is interesting too, but its study and comparison with $\bar{g}$ has been knowingly left for a later time.

[2]Arbitrary finite precisions in the measurements of any coordinates or parameters on which the metric depends can naturally be taken into account in a similar fashion.

Order of magnitude and signs of the apparent energy-density, pressures and scalar curvature. Comparison with the de Sitter vacuum. Let us now discuss the main result of this article, namely expressions (37) for the energy-density and pressure associated with the apparent matter and expression (39) for the corresponding scalar curvature of space-time.

All these quantities clearly tend to zero as R tends to infinity. A rough quantitative estimate of the cumulated effect of the coarse-graining is the ratio ρ of the mass-energy of the apparent matter contained in the volume $R > 2M$ to the mass M of the black hole. In order of magnitudes, one finds :

$$\rho \sim \frac{1}{M} \int_{2M}^{+\infty} \frac{a^2 M^2}{R^6} \, R^2 dR, \tag{40}$$

so that

$$\rho \sim \frac{a^2}{M^2}. \tag{41}$$

Thus, $a \sim M/10$ leads to $\rho \sim 1/100$ whereas $a \sim M$ gives $\rho \sim 1$; naturally, this last typical value of ρ for $a \sim M$ should not be taken too seriously because all the calculations presented in this manuscript were made under the assumption $R \gg a$. If $a \sim M$, our evaluation of the mean metric $\bar{g}$ breaks down for R too close to $2M$ and so does our evaluation of ρ. The calculation nevertheless indicates that a coarse-graining $a \sim M$ will probably produce an apparent matter of mass-energy at least comparable to the mass M of the black hole. The effect will probably be even more important if $a > M$ or $a \gg M$.

Another point deserves further comment. Indeed, the energy density and one of the two pressures of the apparent matter are negative. Imagine now an observer who has access, beyond the coarse-grained metric (26), to a direct evaluation of the mean value $\langle T(\boldsymbol{\omega}) \rangle$ of $T(\boldsymbol{\omega})$, which vanishes identically. This observer may then interpret (37) by associating to the 'vacuum' a non-vanishing energy density and two pressures, a negative one and a positive one. As already pointed out, this brings to mind the recent observational evidence [S$^+$03] for a non-vanishing cosmological constant Λ. Let us now elaborate on this.

The observed cosmological constant is positive. It thus endows the large-scale cosmological vacuum with positive energy-density $\varepsilon_\Lambda = \Lambda$ and a (single) negative pressure[3] p_Λ, which is exactly the opposite of the vacuum energy-density. The cumulated effect of both ε_Λ and p_Λ is best displayed by evaluating two different scalar quantities; the first of these invariants is the associated scalar curvature $\mathcal{R}_\Lambda$ of the cosmological vacuum, defined as the scalar curvature of the 'empty' de Sitter universe with vacuum stress-energy tensor $\mathcal{T}^{\mu\nu} = \Lambda g^{\mu\nu}$; the second scalar $\tilde{\mathcal{R}}_\Lambda$ reflects the 'magnitude' of the so-called trace-free Ricci tensor [PR84] of the same space-time:

$$\tilde{\mathcal{R}}_\Lambda = 8\pi \left[\left(\mathcal{T}_{\mu\nu} - \frac{1}{4}\, \mathcal{T} g_{\mu\nu} \right) \left(\mathcal{T}^{\mu\nu} - \frac{1}{4}\, \mathcal{T} g^{\mu\nu} \right) \right]^{1/2}. \tag{42}$$

[3] The pressure tensor is then isotropic.

A direct calculation gives [HE73] $\mathcal{R}_\Lambda = -4\Lambda \times 8\pi$ and $\tilde{\mathcal{R}}_\Lambda = 0$. In particular, a positive cosmological constant thus induces a *negative* curvature on space-time, which traces the hyperbolic character of the de Sitter expansion.

Unlike ε_Λ, the vacuum energy-density of the coarse-grained Schwarzschild space-time is negative. The pressure tensor of this coarse-grained space-time is anisotropic with two eigen-pressures. One of these eigenpressures is positive, but the other one is negative, as p_Λ. The cumulated effect of these vacuum energy-density and pressures is best compared to the effects of a cosmological term by evaluating the same invariants as those just computed for the de Sitter space-time. Contrarily to $\tilde{\mathcal{R}}_\Lambda$, the 'magnitude' $\tilde{\mathcal{R}}$ of the trace-free Ricci tensor associated with the averaged Scwarzschild space-time is found to be non-vanishing. Indeed, a direct calculation shows that:

$$\tilde{\mathcal{R}} = \frac{18\sqrt{2}}{5}\,\frac{a^2 M^2}{R^6} \tag{43}$$

But *the scalar-curvature $\mathcal{R}$ is, as $\mathcal{R}_\Lambda$, negative* (see (39)). We think this striking result might indicate that at least part of the cosmological vacuum stress-energy may be due to the large-scale averaging of small-scale structures in the Universe. This claim or hypothesis can surely not be proven with the material presented in this article, but our results indicate that this point deserves a more extended investigation.

In the meantime, it is very tempting to try and extrapolate at least the order of magnitudes indicated by our results to a more general astrophysical or cosmological context. This is the purpose of our next paragraph.

Further comments about the possible astrophysical or cosmological implications of our results Let us now extrapolate what has hiterto been presented up to cosmological scales. Our reasoning below is only heuristic and we give the result "as is", hoping to provide at least a loose order of magnitude for the gravitational effects of fluctuations in the large-scale Universe.

The results above suggest that the difference between observing a 'point-like' object of mass M and an object of mass M spread over a characteristic spatial scale (distance) a can lead to a difference in the observed energy-density which scales as:

$$\varepsilon \sim \frac{1}{8\pi}\, G\, \frac{a^2 M^2}{d^6}, \tag{44}$$

d being the 'distance' from the object; here we have abandoned the canonical units and have introduced the gravitational constant G in view of subsequent numerical estimates.

This suggests that the difference between observing a homogeneous object of mass density ρ and spatial size L and a non-homogeneous medium of average density ρ, and size L as well, but having fluctuations of order $\delta\rho$ at characteristic spatial scale a, leads to a difference in energy-density which might behave like:

$$\delta\varepsilon \sim \frac{1}{8\pi}\, G\, \frac{a^2 (\delta\rho\, L^3)^2}{d^6} \tag{45}$$

at distance d from the object. So, following this line of reasoning, the relative variation in energy-density, defined as the ratio of $\delta\varepsilon$ by the average mass-energy ρc^2 of the object, would behave like

$$\frac{\delta\varepsilon}{\rho c^2} \sim \frac{1}{8\pi}\,\frac{G}{c^2}\,\rho\,\frac{a^2\left(\frac{\delta\rho}{\rho}\,L^3\right)^2}{d^6} \tag{46}$$

Let us now boldly apply (46) to the universe itself; this might give some indications on how important the averaging of inhomogeneities might be on the cosmological scale. Let L_U be the Hubble-length and suppose that the characteristic spatial scale of the variations is $a = \alpha L_U$. Suppose also that the observation is made at a distance $d = \lambda L_U$. We get

$$\frac{\delta\varepsilon}{\rho c^2} \sim \frac{1}{8\pi}\,\frac{G\,\rho_U\,L_U^2}{c^2}\,\frac{\alpha^2}{\lambda^6}\left(\frac{\delta\rho}{\rho}\right)^2, \tag{47}$$

where ρ_U stands for the mean density of the Universe.

Both α and $\delta\rho/\rho$ characterize the fluctuations and λ characterizes the distance of observation. On the other hand, $G\rho_U L_U^2/c^2$ does not depend on the fluctuations themselves or on the distance from which they are observed. In the standard cosmological context, this ratio therefore plays the role of a 'fundamental' constant which fixes the order of magnitude of the vacuum stress-energy obtained by averaging a given fluctuation. If one chooses $\rho_U = \rho_{\text{lum}}$, the density of the luminous matter in the universe, one finds, with [KT90] $G = 6.7 \cdot 10^{-8} \cdot \text{cm}^3 \cdot \text{g}^{-1} \cdot \text{s}^{-2}$; $\rho = 10^{-29} \cdot \text{g} \cdot \text{cm}^{-3}$; $L_U = 10^{28}\,\text{cm}$ and $c = 3 \cdot 10^{10} \cdot \text{cm} \cdot \text{s}^{-1}$:

$$\frac{G\,\rho_U\,L_U^2}{c^2} = 0.07. \tag{48}$$

One can also set ρ_U equal to the critical density ρ_{crit}. The critical density is given ([KT90]) by

$$\rho_{\text{crit}} = \frac{3}{8\pi}\,\frac{c^2}{G\,L_U^2} \tag{49}$$

and one then obtains:

$$\frac{G\,\rho_{\text{crit}}\,L_U^2}{c^2} = \frac{3}{8\pi} \approx 0.12 \tag{50}$$

in good agreement with (48), as far as the order of magnitude is concerned. This result practically means that non-linear statistical effects in General Relativity tend to show up precisely at densities around the critical one, which seems quite natural. This indicates that averaging both gravitational fields and energy-densities on a cosmological scale may lead to highly non-trivial and possibly systematic effects whose complete study is however beyond the scope of the present article.

7 Summary and conclusion

This article has been devoted to a first application of the averaging formalism for general relativistic gravitational fields presented in [Deb04b, Deb04a]. We have considered a particular statistical ensemble of space-times which can be physically interpreted as representing a Schwarzschild black hole observed with a finite precision in 'spatial' coordinate measurements. The mean gravitational field associated with this ensemble is *not* a vacuum solution to Einstein's equation. On the contrary, the mean space-time appears as filled with matter; the non-vanishing stress-energy tensor of this apparent matter has been calculated explicitely for points whose 'distance' to the black hole is much larger than the retained coarse-graining. The apparent matter can be characterized by an energy density and two distinct pressures. The energy density and one of the pressures are negative, while the other pressure is positive. The overall effect can be traced by the associated apparent scalar curvature of the vacuum regions, which is negative. This effect brings to mind the negative scalar curvature associated with a positive cosmological constant and this point has been discussed thoroughly; in particular, the above similarity not withstanding, there is naturally a difference between the obtained mean space-time and de Sitter space-time; this difference is reflected by the trace-free Ricci tensor, which vanishes for de Sitter space-time and does not vanish for the mean space-time which describes a Schwarzschild black-hole observed with finite precision.

Moreover, extrapolating the conclusions of this article to a broader astrophysical or cosmological context, we have argued that averaging gravitational fields and energy-densities on a cosmological scale might induce some highly non-trivial and possibly systematic effects, endowing for example the cosmological vacuum with a non-vanishing apparent stress-energy density.

Let us finally mention some of the many possible extensions to this work. One should first of all study systematically the coarse-grained metric obtained in this article. What are the geodesics in this gravitational field? Is there an event- or a Cauchy-horizon? And, if the coarse-grained 'object' qualifies as a black hole, how is the entropy of this coarse-grained black hole related to the entropy of the unaveraged Schwarzschild black hole?

As already mentioned, the same work should also be undertaken on other statistical ensembles of Schwarzschild black holes, associated with a physically different coarse-graining (for example, a coarse graining in Schwarzschild coordinates and not in Kerr-Schild coordinates); and, naturally, one should also evaluate the effects of finite precision measurements on Kerr black holes.

The very general question one would like to answer is: how does an arbitrary, spatially and temporally fluctuating gravitational field appear after coarse-graining? In particular, is the scalar curvature of the mean space-time always lower than the curvature of the unaveraged space-time? And, more precisely, what about the energy-density and pressures of the apparent matter? These questions will probably be best answered numerically. A first step in this direction may be the study of a collection of randomly placed black holes, which would thus serve as a very crude model of 'fluctuating' space-time. We hope to address these questions in subsequent publications. Their importance to astrophysics, cosmology, quantum field theory in curved space-time and quantum gravity can surely not be overestimated.

References

[Buc00] T. Buchert. On average properties of inhomogeneous fluids in general relativity: dust cosmologies. *Gen. Rel. Grav.*, 32:105–26, 2000.

[Buc01] T. Buchert. On average properties of inhomogeneous fluids in general relativity: perfect fluid cosmologies. *Gen. Rel. Grav.*, 33:1381–405, 2001.

[Cha92] S. Chandrasekhar. *The Mathematical Theory of Black-Holes*. Oxford University Press, Oxford, 1992.

[Deb04a] F. Debbasch. Mean field theory and geodesics in general relativity. *Accepted for publication by Eur. Phys.J.B*, 2004.

[Deb04b] F. Debbasch. What is a mean gravitational field? *Eur. Phys. J. B*, 37(2), 2004.

[Fut91] T. Futamase. *Prog. Theor. Phys.*, 86:389–99, 1991.

[Fut93] T. Futamase. *Prog. Theor. Phys.*, 89:581–97, 1993.

[Fut96] T. Futamase. Averaging of a locally inhomogeneous realistic universe. *Phys. Rev. D*, 53:681–9, 1996.

[GS94] G.R. Grimmett and D.R. Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford, 2nd edition, 1994.

[HE73] S.W. Hawking and G.F.R. Ellis. *The large scale structure of space-time*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge, 1973.

[Kas92] M. Kasai. Construction of Inhomogeneous universes which are Friedmann-Lemaitre-Robertson-Walker on average. *Phys. Rev. D*, 69:2330–2, 1992.

[KSMH80] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt. *Exact solutions of Einstein's field equations*. Cambridge Monographs in Mathematical Physics. Cambridge University Press, Cambridge, 1980.

[KT90] E.W. Kolb and M.S. Turner. *The Early Universe*. Frontiers in Physics. Addison-Wesley Publishing Company, Redwood City, 1990.

[LL75] L.D. Landau and E.M. Lifshitz. *The Classical Theory of Fields*. Pergamon Press, Oxford, 4th edition, 1975.

[Nak90] M. Nakahara. *Geometry, Topology and Physics*. Institute of Physics Publishing, Bristol and Philadelphia, 1990.

[Pee93] P.J.E. Peebles. *Principles of Physical Cosmology*. Princeton Series in Physics. Princeton University Press, Princeton, 1993.

[PR84] R. Penrose and W. Rindler. *Spinors and space-time.* Cambridge University Press, Cambridge, 1984.

[S+03] D.N. Spergel et al. First Year Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters. *Astrophys. J. Suppl.*, 148:175, 2003.

[Wal84] R.M. Wald. *General Relativity.* The University of Chicago Press, Chicago, 1984.

[Zal97] R. Zalaletdinov. Averaging problem in general relativity, macroscopic gravity and using Einstein's equation in cosmology. *Bull. Astron. Soc. India*, 25:401–416, 1997.

Habilitation à diriger des recherches

La démarche commune à la plupart des travaux présentés ici est l'étude géométrique d'objets quelconques, typiques, ou irréguliers. La typicité est obtenue par l'utilisation, à un endroit ou à un autre, d'ingrédients aléatoires. Cela peut signifier que l'objet a été obtenu par tirage aléatoire dans sa classe, par perturbation aléatoire autour d'un modèle mieux compris, ou bien qu'un objet est fixé mais suffisamment inconnu pour que le seul point d'attaque consiste à l'observer aléatoirement.

Un des objectifs de cette géométrie « synthétique » ou « robuste » est d'obtenir des arguments qui, lorsqu'ils s'appliquent à un espace, s'appliquent aussi bien à des espaces « proches ». Ces derniers peuvent être, par exemple, des espaces discrets comme un graphe, ou bien des variétés dont la géométrie a un grand nombre de fluctuations ou d'irrégularités à très petite échelle.

Le premier invariant géométrique que l'on rencontre en s'éloignant de l'espace euclidien est la notion de courbure, qui apparaîtra souvent dans ces pages. On peut sommairement diviser son influence en deux continents : celui de la courbure négative (ou majorée), et celui de la courbure positive (ou minorée), qui interviendront tous deux ici.

Cette démarche sera appliquée à trois domaines assez différents. Dans le premier, il s'agit de *groupes aléatoires*, dont le comportement donne des indications sur ce que peut être un groupe quelconque, par opposition aux groupes classiques bien connus. Ici c'est la courbure négative qui domine : les groupes aléatoires sont *hyperboliques*, et beaucoup des propriétés qu'on leur connaît tournent autour de ce fait.

Dans le second, nous adopterons un point de vue géométrique sur les *chaînes de Markov*. On verra en particulier comment utiliser ces dernières pour définir une notion de *courbure (de Ricci)* sur des espaces métriques quelconques, qui permet d'étendre certaines propriétés classiques des variétés de courbure positive, comme la concentration de la mesure.

Enfin, la troisième partie, physique plus que mathématique, traite de *relativité générale* : l'équation d'Einstein lie la courbure au contenu en matière de l'espace-temps, et des fluctuations aléatoires à petite échelle, nulles en moyenne et non observées, peuvent avoir un effet non trivial sur la courbure à grande échelle de l'Univers. Cet effet physique de « matière apparente » est étudié dans différentes situations.

The common idea underlying the various works presented here is a geometric study of generic, typical, or irregular objects. Typicality is achieved through the use of random ingredients at one point or another. The object under scrutiny may have been picked at random in its class, or be a random perturbation of a smoother, more symmetric model, or just be a plain metric space with no particular features, for which random measurements provide the only reasonable approach.

One of the goals of this "coarse" or "robust" geometry is to develop geometric arguments that remain valid when considering objects that are "close" to a given one. The perturbed object might not be regular; typical examples include discrete spaces like graphs, or manifolds with many small fluctuations in their metric.

When departing from Euclidean space, the first geometric invariant encountered is curvature; this notion will pervade our work. Its influence can be broadly divided into two realms: that of positive curvature (or bounded below), and that of negative curvature (or bounded above), which entail very different behaviors. Both will be seen here to some extent.

Three fairly different applications will be used to illustrate these principles. Random groups *will come first. Their behavior hints at what a "generic" group looks like, as opposed to the more classical groups we all learn about. Random groups belong with negatively curved spaces: they are* hyperbolic, *and most of their known features arise from hyperbolicity.*

Next, we will develop a geometric viewpoint on Markov chains, *and see how random walk considerations lead to a notion of* (Ricci) curvature *on arbitrary metric spaces. Several classical properties of positively curved manifolds, such as concentration of measure, extend to this setting.*

A bit of general relativity *will come last; our treatment there will be physical rather than mathematical. The Einstein equation relates the matter content of space-time to its curvature in a non-linear way, and small, unobserved fluctuations of matter may vanish on average, yet have a non-trivial effect on the large-scale curvature and dynamics of the Universe. This physical effect of an emerging "apparent matter" is investigated in a variety of situations.*